Essays in environmental and political economics

Sen, S.

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Essays in Environmental and Political Economics

Suphi Şen

8 September 2014

Essays in Environmental and Political Economics

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op maandag 8 september 2014 om 10.15 uur door

Suphi Şen

geboren op 16 juni 1982 te Kiraz, Turkije.

PROMOTIECOMMISSIE:

PROMOTORES:	prof.dr. Bertrand Melenberg
	prof.dr.ir. Erwin Bulte
OVERIGE LEDEN:	dr. Johann Eyckmans
	dr. Manuel C. Oechslin
	prof.dr. Martin Wagner
	dr. Pavel Cizek
	prof.dr. Reyer Gerlagh

Acknowledgements

First and foremost, I would like to thank my supervisors, Bertrand Melenberg and Erwin Bulte, for their support and guidance. They provided me every bit of guidance, and have always been encouraging and motivating. I also would like to thank Herman Vollebergh. I have learned a lot from our collaboration.

Special thanks to my committee members Johann Eyckmans, Manuel C. Oechslin, Martin Wagner, Pavel Cizek, and Reyer Gerlagh for their time, insightful questions, and helpful comments.

I owe a great debt of gratitude to my professors at ITU. In particular, I thank to Ozgur Kayalica, Benan Zeki Orbay, and Yucel Candemir for their encouragement and support. Special thanks to Umit Senesen for inspiring me to follow an academic career.

I am deeply thankful to my family, Semra, Ismail, and Ozgur, for their love and support. I dedicate this work to my father, Ismail, who passed away last summer, and to Pia who joined the family recently.

Contents

Chapter 1

Introduction

Environmental pollution and income inequality are among the main problems threatening a global sustainable future, and both are strongly intertwined with the unprecedented rise in economic prosperity since the industrial revolution. Both issues lead to similar questions. The first question is: Does economic growth, without any intervention, eventually lead to lower levels of pollution and economic inequality? The second question is: If not, what are the strategies that should be followed? Broadly speaking, this thesis revolves around these questions.

The next section formulates more precisely the questions, which are the topic of this thesis. In Section 1.2, a concise summary of the main hypotheses and the results are presented. Final section of this chapter provides the main implications of the thesis, and elaborates on the relation between environmental pollution and income inequality.

1.1. Questions

Our first quetion is: Should governments intervene to prevent excess increase in pollution or income inequality. If the advantages of economic growth can lead to lower levels of pollution and economic inequality, then there is no need to worry about their consequences on pollution or income inequality. In this case, governments do not face a trade off in optimizing their economic policy by accounting for pollution or income inequality. This idea initiated a substantial literature called the Kuznets Curve following Kuznets (1955) where the focus is on income distribution. One decade later, Grossman and Krueger (1991) suggest the same hypothesis for environmental pollution, and name their hypothesis as Environmental Kuznets Curve (EKC). More precisely, it is hypothesized that although income growth leads to higher pollution and income inequality at the initial stages of economic development, once a turning point is reached, we should see a decline in pollution and income inequality, leading to an inverted-U shaped relation.

Hence any growth enhancing policy is a win-win strategy: While promoting economic growth, the turning points of environmental pollution and income inequality is brought forward.

This thesis starts with an empirical investigation of the Kuznets Curve hypothesis for environmental pollution. Next, the attention is directed to the second question where the focus is again on environmental pollution. The question is that what should be done if there is no Environmental Kuznets Curve. One of the most interesting answers to this question comes from Porter and Van der Linde (1995). Similar to the EKC hypothesis, according to Porter and Van der Linde (1995), there is no trade-off facing the governments in their environmental policies. That is, Porter and Van der Linde (1995) also suggests a win-win strategy. However, in contrast to the EKC hypothesis, Porter argues that more stringent environmental regulations can enhance, not only the environmental outcomes, but also the economic outcomes. The focus in this thesis is on the effect of environmental regulations on aggregate innovation. The question is: How can environmental regulations increase, not just the green, but overall innovation.

The last chapter turns back to the Kuznets Curve, but this time for income inequality. This chapter investigates one of the most important potential factors which might be consequential for a Kuznets Curve in income inequality which is political transitions. According to Acemoglu and Robinson (2000), industrialization in the pre-transition period increases inequality. However, increasing inequality leads to social unrest forcing the elites to adopt democracy where income redistribution targets the poor. This leads to a lower income inequality in the post-transition period. This might suggest that any country following an industrialized growth path might end up in a democratics political system leading to lower inequality levels. Therefore, it is important to understand the forces leading to democratic transition which can be a consequence of, and also consequential for income inequality. This is a task undertaken in the final chapter of this thesis.

1.2. Results

In the EKC literature, it is argued that when income increases beyond a threshold level, environmental degradation will start to decrease. Chapter 2 proposes a new estimation strategy in order to investigate the EKC hypothesis. A fundamental problem in the reduced form EKC estimations is that, specifying a functional form for the time related effects, such as linear or quadratic time trends, is consequential for the estimated shape of the income effects. Therefore income related (scale) effects are not identified. Firstly, following Vollebergh et al. (2009), we apply pairwise differencing strategy in order to identify the scale effects without specifying any functional form for the time related effects. Secondly, our proposed parametric and non-parametric estimation strategies are a combination of recent econometric techniques controlling for cross-sectional dependence, panel non-stationarity, and non-linear transformation of non-stationary covariates. Indeed, applying the first and the second generation panel unit root tests, we find that our series are potentially non-stationary. Our results indicate that, although time related effects (constituting technological and compositional effects among others) of the developed regions are negative, which mitigates environmental degradation, this is not sufficient to create a slow-down in the regional and global level CO_2 emissions.

Chapter 3 forecasts future CO_2 emission pathways by extending the estimation strategy proposed in Chapter 2. In Chapter 3, also the time related effects are estimated by treating them as residual data from the pairwise differencing estimation. The power of the pairwise differencing approach in forecasting future emissions is that a potentially non-linear relation is decomposed into its positive and possible negative components, which enables one to extrapolate these different trends separately. It is shown that, although China's growing income is a strong contributor to the global emissions, the main reason leading to a pessimistic scenario is that the negative trend in the estimated time effects of the developed regions is not sufficiently strong. Therefore, even if the recent high economic growth experienced in China would come to a halt, a slow-down in the increase of global emissions seems to be unlikely.

The fourth chapter investigates the effects of environmental regulations on innovation. It is hypothesized that depending on the distribution of ownership structure of firms in a country, environmental regulations might have an innovation encouraging effect. The hypothesis is motivated with an aggregate R&D model with environmental externalities. It is further assumed that a fraction of firms are controlled by so-called satisficing managers whose only interest is to avoid bankruptcy, while the rest of the firms are governed by owners who maximize their profits. By concentrating on these two extreme corporate governance structures, it is possible to construct a simple country level indicator for the ownership structure, which is the fraction of managerial firms in the economy.The implications of the model are tested by using non-linear count data estimation techniques. The estimation results show that in countries where managerial firms are more prevalent, stricter environmental regulations are more innovation encouraging.

Chapter 5 investigates the potential link between the roles of *elite-poor* and *intra-elite* conflict in democratic transitions. While the former paradigm places revolutionary pressures from low income groups at the center of the analysis, the later paradigm puts forward strategic choices of competing elite factions as a factor leading to democratic transitions. Despite the substantial literature from these two perspectives, the potential relation between the intra-elite and the elite-poor conflict is an untouched area. This chapter puts forward a potential link, arguing that these two potential factors are interrelated. At the center of the analysis, there is the collective action problem of the masses and intra-elite conflict, forcing some elite factions to employ potential de facto power of the masses. It is shown that democratic transitions due to intra-elite conflict are not possible in relatively equal societies. Therefore, the preconditions for a consolidated democracy, put forward by *the elite-poor conflict view*, which is a low income inequality, and by *the intra-elite conflict view*, which is a unified elite structure, are consistent.

The setting in Chapter 5 also allows to analyze the different paths in political revolutions. It is shown that depending on the intra-elite inequality, countries might follow different paths following a revolution. Some revolutions might lead to democracies like the French revolution. On the other hand, some revolutions might lead to more autocratic regimes like in China and Russia in the first half of the twentieth century.

1.3. Main Implications

The results in this thesis, about the relation between economic activity and environmental pollution, calls for strong policy interventions. Chapter 2 suggests that there is no sign of a slow down in the carbon emissions associated with economic growth. The future forecasts of carbon emissions provided in Chapter 3, indicates that global carbon emissions will continue to rise steadily in a business-as-usual scenario. These results indicate the importance of policy intervention in order to achieve environmental goals. The good news is that Chapter 4 shows that the economic costs of environmental regulations might not be very high.

The results of Chapter 5 on political transitions indicate that, not only the elite-poor income inequality, but also the income inequality within the elite might have serious consequences on the post-transition political characteristics of a country. While many political transitions in the history has let to stable democracies where the income inequality might be expected to decrease, many more transitions resulted in stable non-democracies, resulting in a stable and very high income inequality.

Finally, it is important to highlight the relationship between environmental pollution and income inequality as a future research area. A point which remains untouched is that the ones who suffer most from environmental pollution and its possible consequence of a drastic change in climate are those who stay at the bottom of the income distribution. Therefore, a crucial point about the future of environmental policies is regarded with how preferences of low income income segments of a society are translated into policy outcomes through political institutions, and how these political institutions evolve over time.

Chapter 2

The Environmental Kuznets Curve: Identifying Nonlinear Nonstationary Scale Effects

2.1. Introduction

The Environmental Kuznets Curve (EKC hereafter) postulates an inverted U-shaped relationship between pollution and per capita gross domestic production (GDP). That is to say, using the emission level of some pollutant as a proxy, pollution is assumed to follow an increasing pattern up to a certain level of per capita income and once that level, which is called "turning point," is reached, pollution starts to decline. As the initial study testing the EKC hypothesis in a panel setting, Grossman and Krueger (1991) find an N-shaped relation by using a cubic polynomial of income in levels. Following this, the early empirical papers investigating the validity of the EKC hypothesis use various indicators for environmental degradation, employ different functional specifications of income, and analyze many sub-samples of countries or regions. Although there are mixed conclusions for many environmental indicators, most of the studies in the early literature support an inverted U-shaped relation for the air pollutants such as CO_2 and NO_2.[1] In the

[1]For example, Shafik and Bandyopadhyay (1992) use log-linear and log-quadratic specifications in addition to the cubic specification. They find supportive evidence for the EKC hypothesis that CO_2 and NO_2 emissions follow an inverted U-shape pattern with increasing income. Selden and Song (1994) use a dataset mainly including the developed regions, and confirm the EKC hypothesis but with very high turning points. Various indicators of environmental impact are used. For instance, Panayotou (1993) uses SO_2, NO_X, fine particles, and deforestation, Horvath (1997) employs energy use, Komen et al. (1997) use R&D expenditure on environmental protection, and De Bruyn et al. (1998) use sulphur

more recent literature, these findings have been criticized because of employing possibly unsatisfactory econometric techniques (for detailed discussion see, for example, Borghesi 2001, Stern 2004, Muller-Furstenberger and Wagner 2007, and Galeotti et al. 2009). In this paper we focus on two of these econometric issues.[2]

A first criticism is that per capita income and emission series might be non-stationary, and only if the series are cointegrated, the estimations yield reliable results. Otherwise, we might end up with a spurious regression. Perman and Stern (2003) apply some unit root and cointegration tests both for individual series and using panel data. They find that sulfur emissions, GDP per capita, and its square are all I(1). However, results about a cointegrating relationship are ambiguous. In case of no cointegration they performed their estimations with the first differenced variables. In any case, their estimation results do not support the EKC hypothesis. However, this approach is subject to some criticisms. First, the employed panel unit root and panel cointegration tests are so called first generation tests which rely on a very strong assumption of cross sectional independence. Second, as argued by Muller-Furstenberger and Wagner (2007), a non-linear functional specification of a non-stationary exogenous variable requires an appropriate estimation technique and a cointegration test for the hypothesized relation. Wagner (2008) employs both first and second generation unit root tests on a dataset for 100 countries over the period 1950-2000. Results are very dependent on the type of the test chosen. Furthermore, the estimations, which do not account for cross-sectional dependence, confirm the EKC hypothesis, while, by de-factoring the series in order to eliminate the cross sectional dependence, Wagner (2008) finds no significant evidence in favor of the EKC hypothesis.[3]

emission reduction targets. See Stern (1998), for an extensive literature survey.

[2]Another criticism is raised by Taskin and Zaim (2000) towards the trial and error approach of the parametric estimation of the EKC relation where one needs to assume a functional form. See also Millimet et al. (2003). A further issue about the inadequacies of the early empirical studies is highlighted by Dijkgraaf and Vollebergh (2005) who argue that in panel data estimations, the assumption of homogeneity across countries is a very strong one. See also Martinez-Zarzoso and Bengochea-Morancho (2004).

[3]Galeotti et al. (2009) raises another criticism that the employed unit root and cointegration tests do not allow the order of integration to take non-integer values. By applying fractional unit root and cointegration tests in a panel context which allows the order of integration to be non-integer values, they find mixed results towards the EKC hypothesis. However, as mentioned by Galeotti et al. (2009), their method does not account for cross sectional dependence. Additionally, their estimation strategy does

As second criticism Vollebergh et al. (2009) notice that an identification problem arises when separating the effect of a time related independent variable, in our case income, from time effects. It is argued that the restrictions imposed on the time effects may seriously affect the shape of the income-emission relationship. They propose a flexible identification and non-parametric estimation procedure under the assumption that for each region there is at least one other country or region having the same time effect. Their findings for the 24 OECD countries indicate a clear positive income effect for all regions which is not in line with the EKC hypothesis. The time effects are more likely to be inverted U-shaped but not enough to create an inverted U-shaped pattern in total emissions. However, the drawback of their estimation procedure is that it assumes the variables of interest to be stationary. Moreover, they still impose a strong assumption, namely, that for a given country or region there exists another country or region having the same time effect.

The aim of our paper is to deal with both these econometric criticisms at the same time. We first refine the identification strategy proposed by Vollebergh et al. (2009). Instead of starting from the assumption that two selected countries or regions have the same time effect, as Vollebergh et al. (2009) do and from which they identify the income effect, we *define* the income effect of a country or region, *relative to another country or region*, to be what remains (as function of income) after eliminating the common time effect. The pairwise differencing approach of Vollebergh et al. (2009), applied to a country or region and a paired country or region, exactly takes out this common time effect. What remains is then the difference of the two income effects of the two paired countries or regions (with respect to each other). These two income effects are identified and can be estimated fully nonparametrically, without imposing any functional form restriction. The common time effects which are differenced out are allowed to be fully flexible as well.

Each country or region can be coupled with any other country or region, generating in each case a case-specific decomposition of total emissions into an income and a time effect (and a residual idiosyncratic effect), relative to the coupled region. Without additional assumptions this approach reveals that *the* income effect and *the* time effect of a country

not take into account the nonlinear terms of GDP per capita which would require a separate theoretical framework.

or a region cannot be identified. Vollebergh et al. (2009) are able to identify *the* income and *the* time effect of a country only by assuming that for each country the selected paired country is having the same time effects. However, in their application the actual selection is based on a goodness-of-fit criterion. In terms of our interpretation, they do not identify *the* income and *the* time effect of a country, but, instead, they just present the income and time effect of a country *relative to the best fitting country.* We shall proceed under this alternative interpretation.

Given a pair of regions or countries, we perform the pairwise differencing estimations both parametrically and non-parametrically. To deal with the first criticism, and in contrast to Vollebergh et al. (2009), we take into account the non-stationarity properties of the variables. For the parametric estimations, we adopt the estimation strategy "efficient nonstationary nonlinear least squares" (EN-NLS), suggested by Chang et al. (2001). Parametric estimations have the advantage of requiring smaller datasets; however, a non-parametric approach is also desirable by having the advantage of imposing less structure on the income effects. For the univariate case, there are some studies on non-parametric non-stationary regressions (Wang and Phillips, 2009; Karlsen et al., 2007); however, only recently Schienle (2011) developed an estimator for non-parametric non-stationary regressions with many covariates, which fits our pairwise differencing approach. As our final estimation strategy, we use this estimator in our pairwise differencing strategy.[4]

Pairwise differencing can be applied to any country or region together with a paired country or region. In this study we explore the EKC hypothesis by focusing on two large and important regions, namely, "Western Offshoots" (consisting of the Australia, Canada, New Zealand, and the United States, representing a developed region) and China (representing a developing region). We pair Western Offshoots to Western Europe (i.e., the former EU) and we pair China to "Other Asia" (consisting of Japan and the countries of the Middle East).[5] The environmental quality is proxied by CO_2 emission per capita

[4]However, we focus on the special case where the two-dimensional nonstationarity in the GDP per capita levels of the paired regions turn out to be as nonstationary as in the univariate GDP per capita levels. In this special case, Schienle's estimator becomes the Smoothed Backfitting Estimator, see Schienle (2011) for further details.

[5]The underlying data, also used by Melenberg et al. (2011), consists of a balanced panel from nine regions, where the regional division is geographically based and covers the whole world with only a few small country exceptions.

(Marland et al., 2009). Economic activity is proxied by GDP per capita (Maddison, 2009). Using data over the period 1950 to 2006, we do not find evidence towards a slowdown in the income effects of China relative to Other Asia and of Western Offshoots relative to Western Europe, while the negative time effects, if present at all, do not compensate the positive income effects. Hence, given the investigated pairs, there is no evidence supporting the EKC hypothesis, neither when focusing exclusively on the income effect, nor when considering the income effect jointly with the time effect.

We also compare our estimation results with other estimation approaches that can be used under alternative identification strategies, requiring funtional form restrictions. In line with Vollebergh et al. (2009), we find that functional form restrictions as a way to identify the income and time effect plays a crucial role in the shape of the relation between the economic activity and environmental degradation that one finds in an empirical analysis.

The remainder of this chapter is organized as follows. In the next section we discuss our identification strategy and an alternative one based on functional form restrictions. In Section 2.3 we present our dataset and we investigate the stationarity properties of our data. Section 2.4 then describes the estimation strategies. In section 2.5, estimation results, focusing on China and Western Offshoots, are provided. Section 2.6 concludes. The Appendix to this paper, see Sen et al. (2014a), contains background information and additional material.

2.2. Identification Strategies

In the empirical literature, investigating the EKC hypothesis, the general econometric model is as follows:

$$y_{it} = f(x_{it}, i) + \lambda(i, t) + \varepsilon_{it}, \tag{2.1}$$

where i stands for the cross-sectional units, such as countries or regions, and t represents time. The emissions, denoted with y_{it}, is driven by two effects. The first one is the so called income effect which is denoted with f, and which is a function of x_{it}, GDP per capita. Secondly, λ stands for the time effect. Finally, ε_{it} stands for the idiosyncratic error term. In this very general model, both income and time effect can also be a function of some cross-section specific effects. The functional form (2.1) can be motivated by the so-called IPAT-equation (see, for example, Chertow, 2000), i.e., $I = P \times A \times T$, where I stands for the impact (in our case of carbon dioxide emission), P stands for population, A stands for affluence, and T stands for technology. In per capita- and log-terms we get $\log(I/P) = \log(A) + \log(T)$. Translated into equation (2.1), we model $y = \log(I/P)$ by taking $\log(A)$ as a function f of GDP per capita, and $\log(T)$ as a function λ of time, where both f and λ are allowed to be cross section unit-specific.

In order to identify and estimate the hypothesized relationship between emission and income, we apply two identification (and corresponding estimation strategies). Our first, and main, identification strategy ("pairwise differencing") does not impose any additional functional form restrictions (on top of equation (2.1), but instead interprets what can be estimated, using equation (2.1), without additional functional form restrictions, after taking time differences of y_{it} and y_{kt} of two different regions i and k. The second identification strategy ("baseline strategies"), considered for comparison purposes, imposes functional form restrictions. We start by first describing the baseline identification strategies.

To identify f and λ in equation (2.1) in the baseline strategy, we impose the following functional restrictions in (2.1):

$$y_{it} = q(x_{it}, \beta_i) + \tau(t, \pi_i) + \varepsilon_{it}, \tag{2.2}$$

with $f(x_{it}, i) = q(x_{it}, \beta_i)$ for some *known* function q, depending on a vector of unknown parameters β_i, with $\lambda(t, i) = \tau(t, \pi_i)$, for some *known* function τ, depending on a vector of unknown parameters π_i, and where the error term ε_{it} is assumed to satisfy specific stationarity assumptions. Only β_i and π_i need to be estimated. Identification of f and λ is achieved by imposing functional form restrictions. This approach makes sense, in particular, if there would be external information prescribing the functional form restrictions. Otherwise, these functional form restrictions potentially result in misspecified income and time effects.

Our second identification strategy avoids imposing any functional form restrictions (except the specification in (2.1)) by applying pairwise regional time differencing. Formally, consider two regions i and k collected in $c = \{i, k\}$. Then we define $f_c(x_{it}, i)$ and $f_c(x_{kt}, k)$, the region-specific income effects of regions i and k, respectively, given the set of regions c as follows

$$y_{it} = f_c(x_{it}, i) + \lambda_c(t) + \varepsilon_{c,it},$$
$$y_{kt} = f_c(x_{kt}, k) + \lambda_c(t) + \varepsilon_{c,kt},$$

where $\lambda_c(t)$ represents the common time effect, and where $\varepsilon_{c,it}$ and $\varepsilon_{c,kt}$ are the region-specific idiosyncratic error terms. Applying pairwise differencing to these two equations leads to the following equation:

$$y_{it} - y_{kt} = f_c\left(x_{it}, i\right) - f_c\left(x_{kt}, k\right) + \varepsilon_{c,it} - \varepsilon_{c,kt}. \tag{2.3}$$

Assuming $E\left(\varepsilon_{c,it} - \varepsilon_{c,kt} | x_{it}, x_{kt}\right) = 0$, both $f_c\left(\cdot, i\right)$ and $f_c\left(\cdot, k\right)$ can be estimated fully nonparametrically, without imposing additional functional form restrictions. Moreover, because $\lambda_c(t)$ is differenced out, it can be any function of t.

A different coupling, represented by $c' \neq c$, typically will generate a different income effect $f_{c'}(\cdot, i) \neq f_c(\cdot, i)$ and a different time effect $\lambda_{c'}(\cdot) \neq \lambda_c(\cdot)$. This shows that a region's income is only identified relative to another region, as given by the set c.

2.3. Description and Properties of the Data

Our underlying dataset is a balanced panel for all countries, covering the period between 1950 and 2006. CO_2 emission data consists of the sum of emissions from gas, liquid and solid fuels (based on consumption figures), and from gas flaring and cement production (see Boden et al., 1995; Marland et al., 2009). For each type of fuel, data on annual CO_2 emissions result from three aspects: the amount of fuel consumed, the fraction of the fuel that becomes oxidized, and a factor for the carbon content of the fuel. The fuel types incorporated in the calculations are coal, other solid fuels, crude oil, petroleum products, and natural gas. Total energy use and emissions per country are corrected for exports and imports of fuels, as well as for stock changes, international marine bunkers, and non-energy use of fuels, such as chemical feedstock. The estimation of the amounts of CO_2 released through gas flaring are based on the UNSTAT database, supplemented by estimations from DOE/EIA. The estimations of the amounts of CO_2 released from cement manufacturing are based on figures indicating the quantity of manufactured cement, the average calcium oxide content per unit of cement, and a factor to convert the

Table 2.1: Descriptive Statistics

Balanced Panel: N = 9, T = 57, Observations = 513						
Variable	Unit	Mean	Median	St Dev	Min	Max
GDP	bln 1990 $	2266	1358	2297	185	10655
Emission	m tons	498	316	445	18	1832
Population	mln	487	371	330	88	1470
GDP per capita	1990 $	5820	4098	6057	448	29956
Emission per capita	kg	1480	709	1524	39	5607
Log- GDP pc		1.276	1.410	1.012	-0.803	3.400
Log- Emission pc		-0.253	-0.344	1.249	-3.238	1.724

Note: Values of emissions is carbon equivalent.

calcium oxide content into CO_2 equivalents. Data on GDP and population is taken from Maddison (2009). All figures are expressed in 1990 International Geary-Khamis dollars, using purchasing power parities.

We aggregate data on a country by country basis into nine regions: India, China, "Other Asia", Western Europe, Eastern Europe, Former USSR, "Western Offshoots", Africa, and Latin America. In contrast to the division into regions by the Intergovernmental Panel on Climate Change (IPCC), we distinguish explicitly between Eastern Europe and Former USSR, divide the "old" OECD in Western Europe (old EU) and what we indicate as "Western Offshoots" (Australia, Canada, New Zealand, and the United States), while Japan together with the countries of the Middle East are grouped under the name "Other Asia". Finally, we split the IPCC region ALM into Africa and Latin America. In our empirical analysis we focus on two regions in particular: Western Offshoots (to be paired to Western Europe) and China (to be paired with Other Asia).

In Table 2.1, some descriptive statistics for the nine regions are presented. For all variables, it seems that there are no strong outliers. Considering only the mean, median, standard deviation, maximum and minimum values, all variables seem to be right tailed. We shall take logarithms of the per capita variables to correct for the skewness of the level variables.

In Figure 2.1, GDP per capita series of the regions are presented. Three groups can be distinguished. Western Offshoots and Western Europe have the highest income per capita. India and Africa are always at the lowest income group. Other Asia reaches to the middle income group steadily over the period from 1950 to 2006. On the other hand, China achieves this starting from 1985. Eastern Europe, Former USSR, and Latin America stay in the middle income group throughout the period, although Former USSR

Figure 2.1: GDP Per Capita (thousand US $ - 1990)

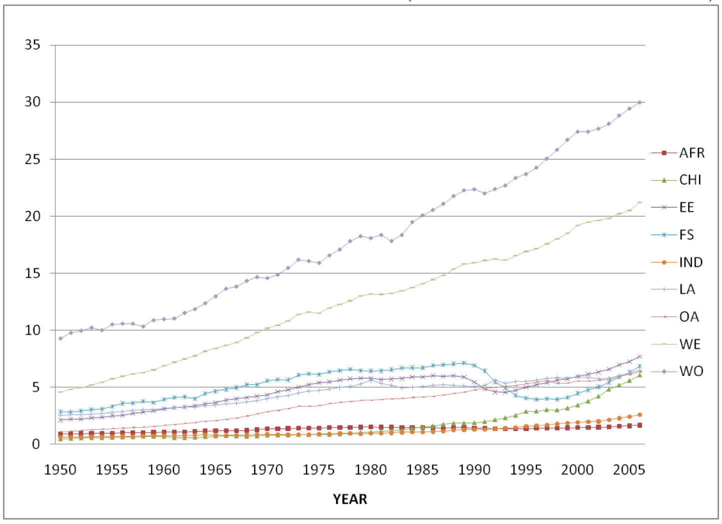
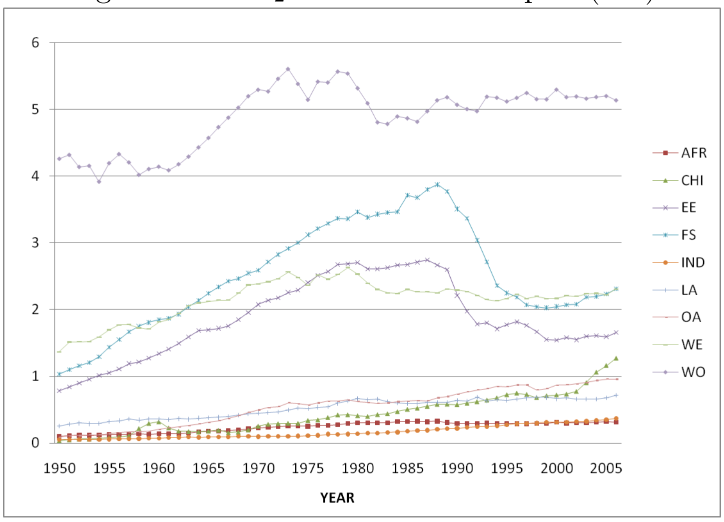


Figure 2.2: CO_2 Emission Per Capita (ton)



experiences a decline following its collapse in 1990.

Figure 2.2 illustrates the corresponding CO_2 emissions per capita. Compared to Figure 1, there are some clear differences. Firstly, Former USSR and China are very close to the high income group in their CO_2 emissions. Secondly, the more rapid GDP rise in the high income countries observed in Figure 1 is not observed in the emission series. Changes in emission per capita seem to be more similar across regions when compared with changes in GDP per capita. Lastly, while GDP per capita series seem to increase in time, the high emission countries seem to experience a decline in their CO_2 emission at a point in time; however, for low emission countries this is not so clear. This could be evidence towards the EKC hypothesis.

Next, we turn to the stationarity properties of our variables, considering our dataset as a panel including all nine regions. To test the stationarity of our variables in levels, we focus on the second generation panel data unit root tests, that take cross sectional correlation into account. Here, we focus on Bai and Ng (2004). In Sen et al. (2014a) we also report the outcomes of other second generation unit root tests.[6]

Bai and Ng (2004) consider a multi-factor framework called "Panel Analysis of Nonstationarity in Idiosyncratic and Common Components" (PANIC) where the factors and idiosyncratic components are analyzed separately and hence allow for cross-unit cointegration. Furthermore, this method allows testing the number of factors with a unit root. Their model is as follows:

$$\begin{aligned} z_{it} &= d_{it} + \lambda_i' F_t + E_{it} \\ F_t &= F_{t-1} + \eta_t \\ E_{it} &= \rho_i E_{it} + e_{it} \end{aligned}$$

where z_{it} is the variable under consideration, with index i referring to region i and index t to year t, where F_t is the vector of common factors, E_{it} is the idiosyncratic component, and d_{it} is the deterministic component of the data generating process which indicates whether the model includes a constant or a trend. The disturbances η_t and e_{it} are assumed to be white noise processes.[7]

In this data generating process, one or more of the common factors might follow a random walk. Therefore, the standard factor analysis does not apply to identify the factor loadings. To deal with this issue, Bai and Ng (2004) suggest basing the principal component analysis on the first differences of the series. The standard PANIC analysis uses the selection criteria suggested by Bai and Ng (2002) to determine the number of common factors. However, this criterion performs poorly when the cross-sectional dimension is small, like in our case. So, in order apply these tests, we assume the number of common factors to be at most three.

[6]For the sake of completeness we also present results of univariate unit root tests and first generation panel data unit root tests in the Appendix to this paper, see Sen et al. (2014a). However, when investigating cross sectional dependence in Sen et al. (2014a), we clearly find evidence for cross sectional dependence (both via common factors and via idiosyncratic components).

[7]The r-dimensional disturbance term η_t is modeled as $\eta_t = C(L)u_t$, with $C(L) = \sum_{j=0}^{\infty} C_j L^j$ and u_t i.i.d., where $\text{rank}(C(1)) = r_1 \in [0, r]$, with r_1 number of I(1) factors and $r - r_1$ the number of stationary factors.

Table 2.2: Bai & NG (2004) PANIC Results

	Number of Common Factors	Number of Common Stochastic Trends			Idiosyncratic Components	
		ADF	MQ-c	MQ-f	BN_N	$BN_{\chi 2}$
	Model with individual intercept					
log emissions pc	1	1	na	na	0.977	0.996
	2	2	2	2	0.076	0.087
	3	3	3	3	0.000	0.000
log GDP pc	1	1	na	na	0.529	0.485
	2	2	2	2	0.642	0.606
	3	3	3	3	0.492	0.448
	Model with individual intercept and trend					
log emissions pc	1	0	na	na	0.998	1.000
	2	1	2	2	0.910	0.946
	3	2	3	3	0.731	0.708
log GDP pc	1	1	na	na	0.990	1.000
	2	2	2	2	0.991	1.000
	3	3	3	3	0.987	0.999

Note: Unit root tests on the common factors are conducted at 5% significance level. For the idiosyncratic components, presented values are the p-values of the pooled tests. Abbreviation "na" indicates the test is not applicable.

In Table 2.2, the results of the Bai and Ng test are presented. The first column indicates the given number of common factors which is assumed to be one, two, or three. In order to investigate the number of common factors with a unit root, the Bai and Ng (2004) method applies ADF unit root tests, labeled as "ADF" in Table 2.2. In case of more than one common factor, individual ADF tests may over-state the number of common stochastic trends (Bai and Ng, 2004), since only the space spanned by the factors can be estimated. Therefore, Bai and Ng (2004) suggests two tests (MQ-f and MQ-c), which are slightly modified versions of the cointegration tests suggested by Stock and Watson (1998). The null hypothesis states the number of common stochastic trends against the alternative that it is less than the stated number in the null hypothesis. The test is applied successively by decreasing the number of stochastic trends in the null hypothesis as long as it is rejected.

For the idiosyncratic components, Bai and Ng (2004) provide two test statistics, one that is asymptotically normally distributed (BN_N) and the other one that has an asymptotic chi-square distribution (BN_{ξ^2}). Both tests depend on pooling the p-values from the ADF tests applied to individual idiosyncratic components

2.4. Estimation Strategies

The results in Table 2.2 indicate that there are likely multiple common factors with a unit root, while also the idiosyncratic components of both the GDP and emission series seem to have a unit root, at least, when allowing for a deterministic trend. Given these outcomes, we shall proceed under the assumption that both the GDP and emission series are nonstationary, in line with the findings of Wagner (2008) who also applied the PANIC analysis (and other tests) to carbon dioxide and GDP per capita, using data over 100 countries during the period 1950 to 2000.

In this section we discuss our estimation strategies. We start by first describing the baseline estimation strategies to estimate (2.2). Next, we discuss the estimation strategies to estimate (2.3). In both cases the estimation strategies take the non-stationarity of our variables into account.

2.4.1. Baseline Strategies: Functional Form Restrictions

To deal with the presence of the nonlinear transformations of the nonstationary regressors in (2.2), which requires a different asymptotic theory than the usual nonlinear least squares, we use the "efficient nonstationary nonlinear least squares" (EN-NLS) estimator of Chang et al. (2001).[8]

Following Chang et al. (2001), we first estimate equation (2.2), using standard Nonlinear Least Squares, to get $\hat{\varepsilon}_{it}$. Secondly, for $\nu_t = \Delta x_t$ we run the following auxiliary regression:

$$\nu_t = \hat{\Pi}_1 \nu_{t-1} + \hat{\Pi}_2 \nu_{t-2} + \cdots + \hat{u}_t,$$

where the lag number of ν is determined by the criterion given by Chang et al. (2001). Now, we are ready to transform the dependent variable in (2.2) to obtain the EN-NLS. Indicating the transformed variable with a star, the transformed dependent variable is given as:

$$y^*_{it} = y_{it} - \hat{\sigma}_{\varepsilon u} \Sigma^{-1}_{uu} \hat{u}_{t+1},$$

where $\hat{\sigma}_{\varepsilon u} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{i,t+1} \hat{u}'_{it}$ and $\hat{\Sigma}_{uu} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{i,t} \hat{u}'_{i,t}$. We are then able to estimate

[8] EN-NLS estimation also allows to incorporate a linear trend and stationary regressors, in addition to multiple I(1) regressors. In our case, we do not have stationary variables, but only one nonstationary variable and its nonlinear transformation as regressors.

the parameters of interest efficiently, by using the transformed dependent variable in our regression as follows:

$$y_{it}^{*} = q(x_{it}, \beta_i) + \tau\,(t, \pi_i) + \varepsilon_{it}^{*}. \tag{2.4}$$

For both q and τ we choose polynomials, i.e., we specify (2.4) as

$$y_{it}^{*} = \alpha + \sum_{j=1}^{J} \beta_{ij} x_{it}^{j} + \sum_{k=1}^{K} \pi_{ik} t^{k} + \varepsilon_{it}^{*}, \tag{2.5}$$

with α the constant term in this regression equation. Chang et al. (2001) provide further details, including regularity conditions and the quantification of the sampling inaccuracy.

In equation (2.5) the cross correlation is captured by polynomials in x_{it} and t (with ε_{it}^{*} assumed to be uncorrelated over i). Alternatively, to deal with the cross-sectional dependence problem, Pesaran (2006) suggests the Common Correlated Effects (CCE) estimation. In this method, the regression is augmented by cross-sectional averages of both the dependent and independent variables, resulting in

$$y_{it} = \alpha + \sum_{j=1}^{J} \beta_{ij} x_{it}^{j} + b_i \bar{y}_t + \sum_{k=1}^{K} d_{ik} \bar{x}_{t}^{j} + \varepsilon_{it}, \tag{2.6}$$

where $\bar{y}_t$ is the cross-sectional averages of the emission series and $\bar{x}_t$ is the cross-sectional average of the GDP series, and where b_i and d_{ik}, for $k = 1, \cdots, K$ are the corresponding regression coefficients. The underlying logic of the CCE estimation is to proxy the common factors in the error structure, which creates the cross-sectional dependence, by means of the cross-sectional averages of the variables. Thus, the variables of interest are defactored. Kapetanios et al. (2011) show that the CCE estimation also accounts for a multifactor structure, and allows the common factors to be I(1) processes. It is important that CCE does not require the number of factors to be estimated. It is only assumed that the number of unobserved factors remains fixed as the sample size increases. Therefore, if the common factors are responsible for the non-stationarity, equation (2.6) can be estimated without requiring a cointegrating relationship, even if the original variables are non-stationary.

We complement the baseline estimation strategies by also using EN-NLS applied to demeaned variables, as an informal way to mitigate the effects of cross sectional dependence.

2.4.2. Estimation via Pairwise Differencing

In this subsection we describe our estimation strategies for equation (2.3). Next to nonparametric estimation of $f_c(\cdot, i)$ we shall also consider parametric alternatives. We start with the latter.

First, when specializing $f_c(x_{it}, i)$ as in (2.5), and assuming $\varepsilon_{c,it} - \varepsilon_{c,kt}$ to be stationary, we can estimate this regression with "dynamic ordinary least squares" (DOLS), suggested by Saikkonen (1991), which deals with the efficiency problems of the OLS estimation in a cointegration relationship. In DOLS, the regression is augmented by lags and leads of the first differenced regressors. Let $z_{it} = \left(x_{it}, x_{it}^2, \cdots, x_{it}^J\right)'$ and $\beta_i = (\beta_{i1} \cdots, \beta_{iJ})'$, then we have (suppressing the dependence of the parameters on c):

$$y_{it} - y_{kt} = \beta_i' z_{it} - \beta_k' z_{kt} + \sum_{\ell=-q}^{p} (\gamma_{i\ell}' \Delta z_{i,t-\ell} - \gamma_{k\ell}' \Delta z_{k,t-\ell}) + (\varepsilon_{c,it} - \varepsilon_{c,kt}), \tag{2.7}$$

where the terms $\Delta z_{i,t-\ell}$ and $\Delta z_{k,t-\ell}$ are the lagged ($\ell > 0$) or lead ($\ell < 0$) values of the first differenced regressors by means of which DOLS deals with the efficiency problem, and where $\gamma_{i\ell}$ and $\gamma_{k\ell}$ are the corresponding J-dimensional parameter vectors.

A practical problem is that, as reported in Kao and Chiang (2001), the parameter estimates might change substantially with the chosen number of lags and leads. There are several strategies in the literature in order to deal with this problem (see Westerlund, 2005; Kejriwal and Perron, 2008; Choi and Kurozumi, 2012). We adopt the strategy by Choi and Kurozumi (2012), who propose a data dependent choice of the maximum numbers of lags and leads, and the number of lags and leads is chosen based on the BIC.[9]

As alternatives, we use EN-NLS and a nonparametric estimator. Non-parametric estimation techniques have the advantage of putting minimum restrictions on the hypothesized functional relation, which is a very suitable property for our case. Recently, Schienle (2011) shows how to generalize the nonparametric smooth backfitting estimation

[9] In selecting a model specification, using Schwarz Bayesian Information Criterion (BIC) is a frequently referred method. Another criterion is the Akaike information criteria (AIC). Both AIC and BIC use a penalty for the increasing number of regressors. That is, they measure the tradeoff between parsimony and goodness of fit. These two criterion differ in the degree of the penalty for the increasing number of parameters. However, AIC is likely to choose asymptotically overparametrized models, while BIC always selects the true model (Verbeek, 2004, p.285). Therefore, we prefer BIC when the two criteria choose different models.

for additive models suggested by Mammen et al. (1999) to account for non-stationary regressions with many covariates. We focus on the special case where the two-dimensional nonstationarity in the regressors of the paired regions is as nonstationary as in the univariate regressors. For the paired GDP-s per capita this seems to be the case. If so, Schienle's generalized estimator is the smoothed backfitting estimator, see Schienle (2011) for further details. An accessible description of the smoothed backfitting estimator including its computation in a practical application can be found in Nielsen and Sperlich (2005).[10]

In pairwise differencing we need to pair two regions. For the purpose of comparison, we use the pairs in Melenberg et al. (2011), who propose the "Goodness-of-Fit (GoF) prior" in choosing the pairs. GoF prior chooses the pair for a region among all candidates, such that pairwise differencing estimation gives the lowest sum of squared errors. This means that we couple China to "Other Asia" and Western Offshoots to Western Europe.

In pairwise differencing estimations, it is possible to calibrate the time effects. From equation (2.3), we obtain the emissions depending on the GDP per capita of the region i as $\widehat{f_c}\,(x_{it}, i)$ and region k as $\widehat{f_c}\,(x_{kt}, k)$, where $\widehat{f_c}$ is the estimated f_c from the pairwise differencing estimations. The time effect of each region is constructed by subtracting the income effects from the observed emissions as $\hat{d}_c(i,t) = y_{it} - \widehat{f_c}\,(x_{it}, i)$ for region i, and $\hat{d}_c(k,t) = y_{k,t} - \widehat{f_c}\,(x_{kt}, k)$ for region k. The time effects are homogeneous across paired regions. Therefore, in order to calibrate the time effects, we average these two time effects.[11]

Except for Schienle (2011) (and possibly the CCE-estimation), the estimations require a cointegration relationship, i.e., the residuals in the regression equations should satisfy specific stationarity assumptions. We complement our estimation results by an extensive cointegration analysis (presented in the Appendix). Since the available and implemented cointegration tests do not necessarily exactly match the estimation specifications used in this section, this cointegraton analysis is just to investigate whether cointegration

[10]The only drawback is that there is no bandwidth selection procedure theoretically developed yet, although there is one for the stationary case by Mammen and Park (2005)). Therefore, we prefer to use the common rule of thumb in order to determine the bandwidths. We prefer to use the common rule of thumb, $h = 1.06\hat{\sigma}n^{-1/5}$, in order to determine the bandwidths.

[11]In the companion paper Melenberg et al. (2014), we model these calibrated time effects as a function of time to generate forecasts of future carbon dioxide emissions.

relationship close to our estimation specifications are likely or not.[12]

2.5. Estimation Results

In this section we present the estimation results. We focus on China (paired to Other Asia) and Western Offshoots (paired to Western Europe).[13]

Baseline Estimations—We start with the baseline estimations, using EN-NLS and CCE as the main estimation strategies without pairwise differencing. In Figure 2.3, EN-NLS estimations for China are presented in the first three rows, where the functional form of the individual deterministic trend is different for each row as indicated in the figure. The corresponding estimation results can be found in Table 2.4. The number, "(#)" on the bottom-right corner of each graph indicates the rank of preference by BIC, and the curves for which the highest order term of the polynomial is not significant is indicated with "ns." The only estimation among these three, supporting the EKC hypothesis (in terms of the income effect), is the quadratic equation with no trend. However, it is the less preferred specification by the BIC criterion. As expected for the China case, the rest of the EN-NLS estimations reject the EKC hypothesis. A more important result is that the estimated shape of the income-emission relation for a given polynomial specification changes substantially with the assumed functional form of the deterministic trends. This result reveals the importance to put minimum restrictions on the time effects in order to estimate the shape of the income-emission relation.[14]

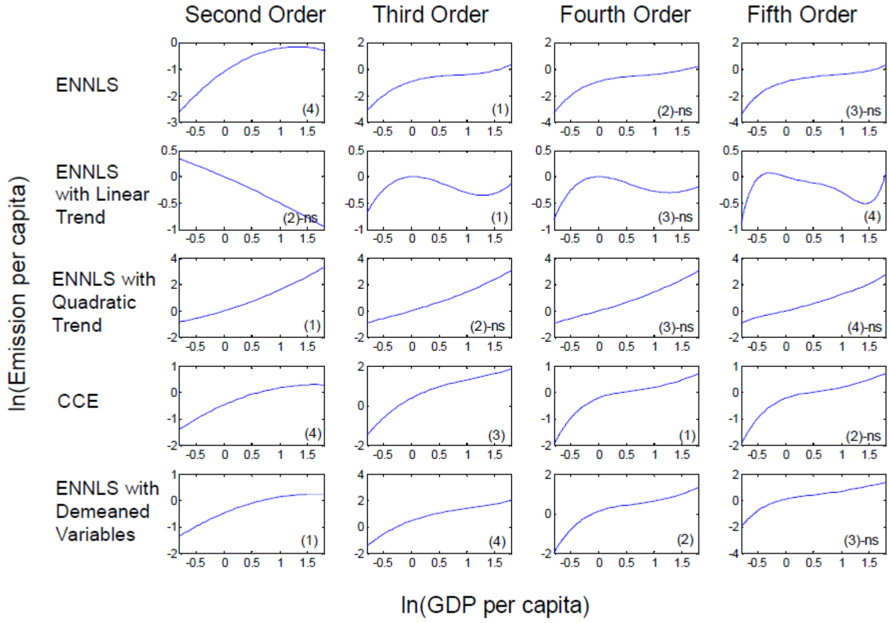
We proceed with the baseline estimations for China with the strategies controlling for common stochastic trends, namely CCE and EN-NLS estimation, with demeaned variables. These are presented in the last two rows of Figure 2.3 (with estimation results in Table 2.4). The estimated curves are very similar to those estimated by the EN-NLS without deterministic trends. This result further highlights the importance of specifying

[12]The error terms in the Schienle (2011) approach also have to satisfy specific conditions. However, we are not aware of a formal test to test these conditions.

[13]The Appendix contains additional estimation results, in particular, under the assumption of homogeneity. In addition, the Appendix also presents some estimation results of the other regions.

[14]This is also confirmed by the estimation results under the assumption of homogeneity, see the Appendix.

Figure 2.3: Baseline Estimations for China

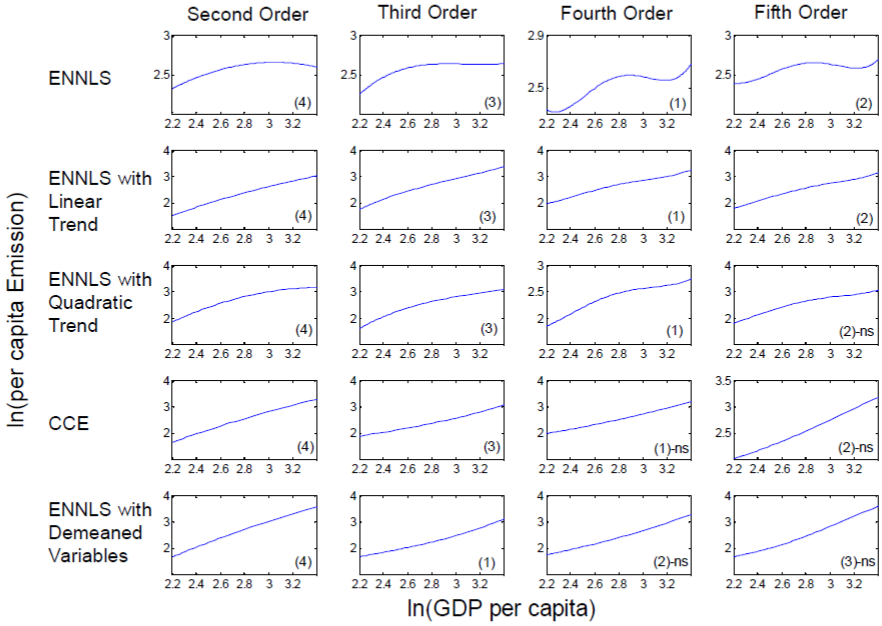


the right functional form for the time effects.

The results for the baseline estimations for Western Offshoots are presented in Figure 2.4 and Table 2.5. None of the estimations predicts an inverted U-shaped curve for the Western Offshoots. Even for this economically most developed region (together with Western Europe), where a downturn in the emissions is expected to be more likely, we find as result that the emissions seem to be rising with increasing income. This is in line with the perspective that the income effect is a scale effect, and, therefore, a positive effect.

Figure 2.5 illustrates the estimated income and time effects for China and Western Offshoots by the EN-NLS estimation, without differencing, with the chosen polynomial specification by the BIC (among all the EN-NLS estimations). For China, this is the quadratic specification for both the income and the time effects. For Western Offshoots, it is the fourth order polynomial with linear trend. The time effects are negative for Western Offshoots which makes sense, since in rich regions one expects to see a decline in the emissions due to technological progress, creating an environment-friendly industry and due to the changes in the sectoral composition by a shift away from the pollution intensive sectors. The overall effect is positive for most of the periods, implying that the time effects are not sufficient to offset the positive income effect. Hence, there is no sign of a slowdown in the total emissions. However, there is a counterintuitive pattern for

Figure 2.4: Baseline Estimations for Western Offshoots



the time effects of China which is inverted U-shaped. This makes the pattern of income and time effects for China very similar to the one for Western Offshoots, i.e., a rising income effect, a declining time effect, and a rising total effect, dominated by the income effect. As discussed previously, this could be a misleading result due to the restrictions on the time effect. As a remedy to this problems, the pairwise differencing approach, which puts the minimum restrictions on the time effects, is applied.

Parametric Pairwise Differencing Estimations—We continue by applying the parametric pairwise differencing strategy, with China paired to Other Asia, and Western Offshoots to Western Europe. The pairwise differencing estimation does not identify the levels of these curves. Therefore, we normalize the curves such that the level of the sample average for each curve is equal to the average level of the observed emission in that region. The results are presented in Figure 2.6 and Table 2.6 for China, and in Figure 2.7 and Table 2.7 for Western Offshoots. Firstly, for both China and Western Offshoots, applying EN-NLS, which controls for the non-stationarity and nonlinear transformations of the income variable, does not change the results over the simple OLS estimation. Secondly, the estimated curves are not in line with the EKC hypothesis (in terms of the income effect), supporting the baseline estimations.[15]

[15]There is a striking difference between the estimations under homogeneity, see Appendix, and heterogeneity. The supportive EKC evidence in the panel estimations seems to be driven by the assumption

Figure 2.5: Baseline - EN-NLS Estimation with Deterministic Trend

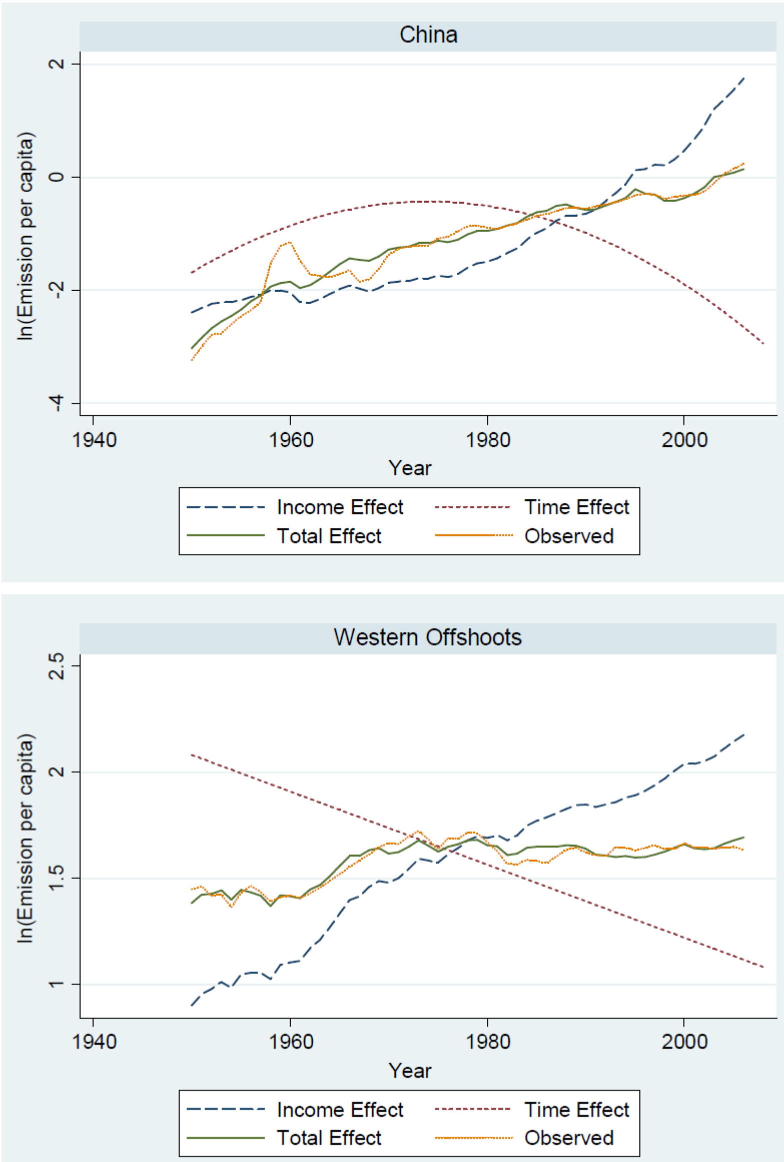


Figure 2.6: Pairwise Differencing Estimations for China

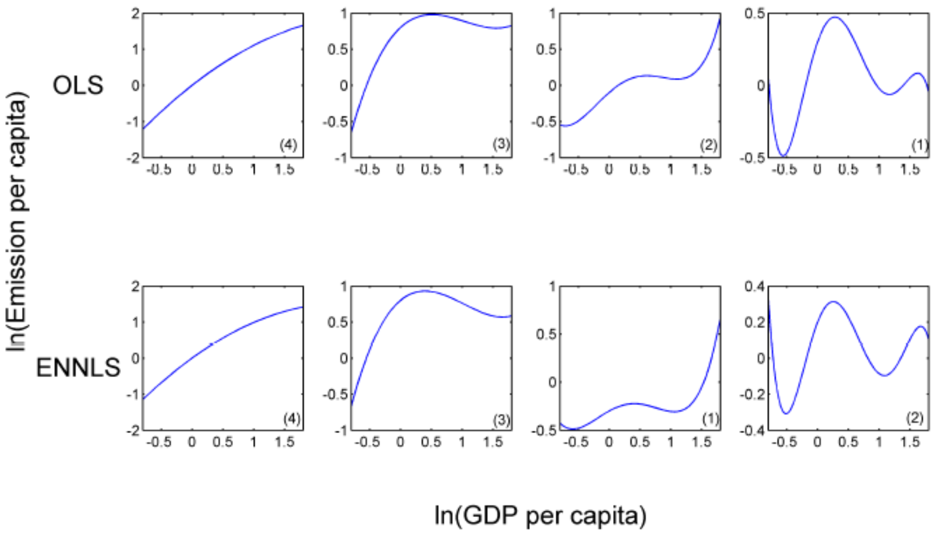


Figure 2.7: Pairwise Differencing Estimations for WO

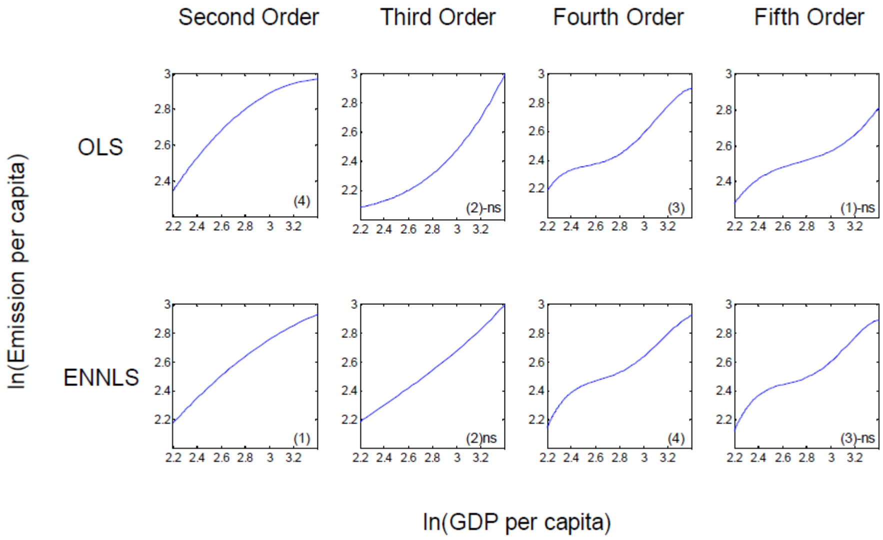
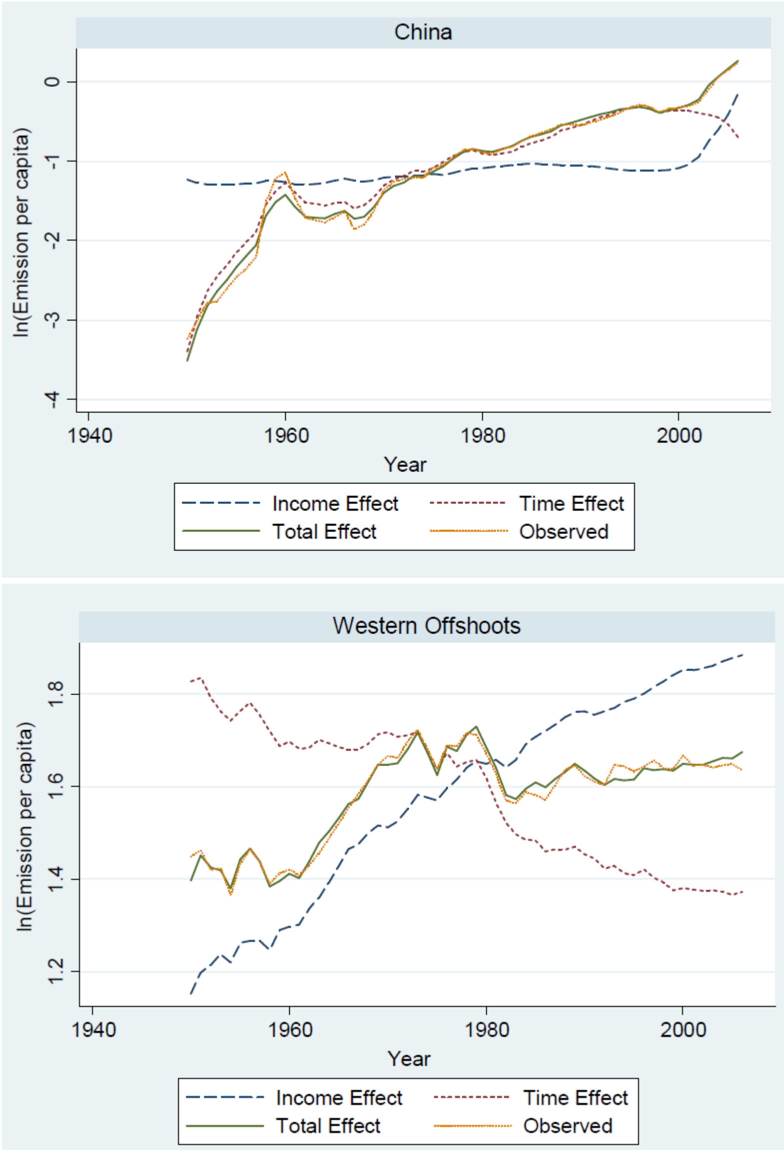


Figure 2.8: Pairwise Differencing – EN-NLS Estimation



In Figure 2.8, for China and Western Offshoots, the estimated income and time effects by the pairwise differencing approach with the EN-NLS strategy is illustrated for the chosen polynomial specification by the BIC. The picture for Western Offshoots remains the same as in Figure 2.5, while for China it is quite different. Specifically, both the time and income effects are increasing, and the time effects are stronger than the income effects. That is, technological and sectoral composition play a major role in China in increasing emissions.

Nonparametric Pairwise Differencing Estimations—Our last estimation strategy is incorporating the non-parametric non-stationary estimator (generalized smooth back-

of homogeneous income effects.

Table 2.3: Cointegration Tests GDP-s Paired Regions

	Western Offshoots - Western EU.	China - Other Asia
ADF Statistics	-3.524	-3.817
MacKinnon p-value	0.036	0.016

fitting) suggested by Schienle (2011) in our pairwise differencing approach. This approach does not require to make a priori functional assumptions for the hypothesized relationship.[16] Also, as the EN-NLS approach, this estimator accounts for the nonstationarity in our variables. Moreover, it fits to the pairwise differencing approach with its additive formulation, and by allowing more than one covariate. Table 2.3 presents cointegration tests for the GPD-s per capita in the paired regions, suggesting that there is a cointegrating relationship between the paired GDP-s per capita. This motivates our choice to specialize the Schienle (2011) generalized smooth backfitting estimator to the Mammen et al. (1999) smooth backfitting estimator, following Nielsen and Sperlich (2005) in the implementation.

The results illustrated in Figure 2.9 mainly support the results of EN-NLS pairwise differencing estimations. For China, the estimated income and time effects are positive, and the time effects are stronger than the income effects. This is the same conclusion with the results of the EN-NLS pairwise differencing estimation. For the Western Offshoots, the income effect is positive, supporting the result of the EN-NLS estimation. The only difference is that for Western Offshoots the time effects seem to be increasing until the 1970s and decreasing afterwards, while the EN-NLS strategy estimates a decreasing time effect throughout the sample period. These results show that once we allow the time effects to be fully flexible, by introducing the pairwise differencing approach and by accounting for nonstationarity and nonlinearity, the estimated patterns for the income and time effects, as well as their implications for the EKC hypothesis, remain robust to using parametric or nonparametric techniques in order to estimate the income effects. The pairwise differencing approach both with the parametric EN-NLS estimator and the nonparametric smooth backfitting estimator of Schienle (2011) does not support the EKC hypothesis.

[16] In the parametric approach, we overcome this problem by estimating polynomials up to fifth degree.

Figure 2.9: Non-parametric Non-stationary Estimation by Pairwise Differencing

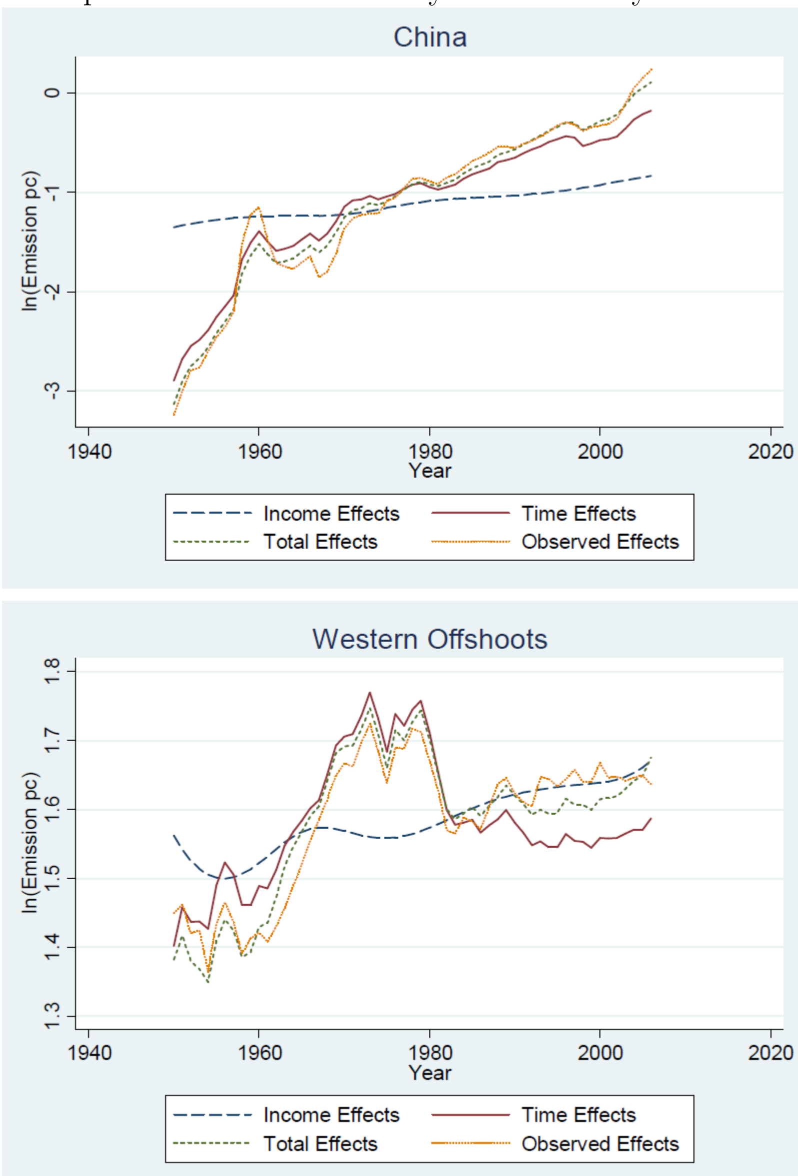
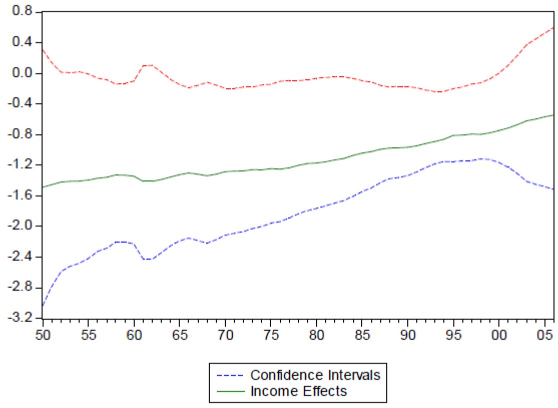


Figure 2.10: Confidence Intervals for China

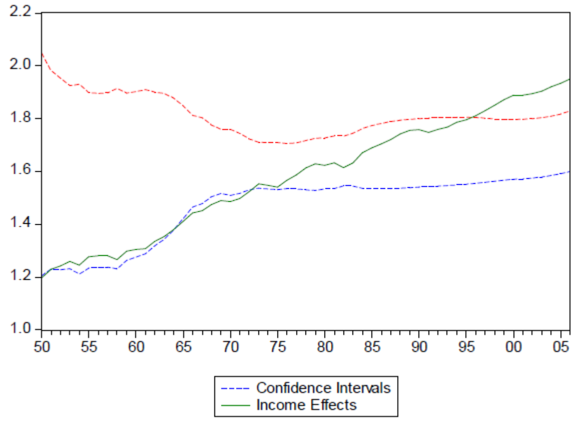


Parametric versus Nonparametric Pairwise Differencing—While the non-parametric approach is fully flexible in the specification of the income effects, it is not as efficient as the parametric approach. For example, non-parametric estimations may suffer from end-of-sample biases, and out-of-sample predictions may be driven by this problem. Therefore, one may prefer parametric estimations in making out-of-sample predictions. However, compared to non-parametric estimation strategies, parametric estimations have the disadavantage of imposing more structure on the income effects. Therefore, it makes sense to check the in-sample performance of the parametric estimations. One way of doing this is to check if our parametric estimations stay inside the non-parametric confidence intervals.

In Figures 2.10 and 2.11, we present the non-stationary parametric estimates of the income effects and the corresponding non-parametric non-stationary 90% confidence intervals for China and Western Offshoots. In constructing the confidence intervals, we follow Nielsen and Sperlich (2005) (see the Appendix for details). For China, the estimated income effects clearly stay inside the confidence intervals, which alleviates the concerns about the structure imposed by the parametric specification. For Western Offshoots the picture is less clear. The estimated income effects do not fully fit inside the confidence intervals. However, the estimated positive pattern for the income effects does not conflict with the non-parametric intervals. This may raise a concern that the estimated pattern by the parametric specification may exaggerate the positive pattern.

Cointegration—In the Appendix we complement the estimation results of this section by corresponding cointegration tests, although these tests do not necessarily match the estimation specifications employed in this section. Our cointegration analysis in partic-

Figure 2.11: Confidence Intervals for Western Offshoots



ular provides some evidence of a cointegrating relationship for the pairwise differencing regressions.

2.6. Conclusion

In this paper we deal with two econometric issues related to the traditional quantification and estimation of Environmental Kuznets Curves (EKCs), namely the lack of identification and the need to use estimation techniques that can handle non-stationary data. To deal with these two criticisms simultaneously, we use pairwise differencing to identify the income effect of a region relative to some other region and we apply nonlinear-nonstationary parametric and non-parametric estimation techniques to estimate the pairwise differenced regressions. Using estimation procedures suitable for non-stationarity is important, since, based on the PANIC-approach proposed by Bai and Ng (2004), we find strong evidence that carbon dioxide emissions and GDP per capita are nonstationary, in line with the earlier literature.

We find that the estimated patterns for the income and time effects are quite sensitive to the specification of the income and time effects (such as linear versus quadratic, or deterministic versus stochastic). This finding indicates the importance of being flexible in the identification, avoiding functional form restrictions. Our pairwise differencing estimations that allow for this flexibility do not support the EKC hypothesis. We do not find clear inverted U-shaped relations, in particular not for an economically developed region like Western Offshoots (relative to Western Europe).[17]

The pairwise differencing approach identifies the income effect of a region relative to another region, allowing consistent estimation. However, the time effects are only calibrated. A natural next step is to construct and estimate a model for the time effects, using these calibrated time effects. This is a topic that we investigate in the companion paper Melenberg et al. (2014).

2.A. Appendix

2.A.1. Introduction

This Appendix contains background as well as additional material. Section 2.A.2 contains the results of univariate unit root tests. Section 2.A.3 presents the outcomes of first generation panel unit root tests. In Section 2.A.4 we test for the presence of cross sectional dependence. Section 2.A.5 contains the results of the second generation panel unit root tests (not included in the main text). Section 2.A.6 shows the outcomes of the unit root tests relevant for the models with pairwise differencing. Section 2.B presents the estimation results under the homogeneity assumption. Section 2.B.1 contains the results of the cointegration tests. Section 2.B.2 presents the construction of the nonparametric confidence intervals, presented in the main text. Finally, Section 2.B.3 contains some estimation results of the other regions.

2.A.2. Univariate Unit Root Tests

In this section we apply the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) univariate unit root tests. Results may be useful in interpreting the results of the panel unit root tests.

Results are presented in Table 2.8 for the logarithm of the emission per-capita and in Table 2.9 for the logarithm of GDP per-capita. The ADF and KPSS tests mostly give conflicting findings; however, for every series, at least one of the tests indicates a unit root. Another crucial point for the following sections is that according to the ADF

[17] This also applies to the other regions, but the estimations under homogeneity support the EKC hypothesis, see Appendix.

Table 2.4: Baseline Estimations for China

PANEL A: ENNLS								
LGDPPC	1.438	***	1.310	***	1.168	***	1.105	***
LGDPPC2	-0.548	***	-1.369	***	-1.378	***	-0.976	**
LGDPPC3			0.573	**	0.933	**	1.071	**
LGDPPC4					-0.188		-0.722	
LGDPPC5							0.221	
Constant	-1.107	***	-0.938	***	-0.941	***	-0.967	***
AIC	21.187		2.646		3.623		5.134	
BIC	27.209		10.675		13.660		17.178	
PANEL B: ENNLS with Linear Trends								
LGDPPC	-0.455		0.052		0.005		-0.329	
LGDPPC2	-0.043		-0.719	***	-0.740	***	0.057	
LGDPPC3			0.361	***	0.559		0.845	**
LGDPPC4					-0.101		-1.415	*
LGDPPC5							0.550	*
Year	0.068	***	0.047	***	0.046	**	0.053	***
Constant	-135.044	***	-93.208	***	-91.068	***	-104.948	***
AIC	-5.536		-10.098		-8.294		-10.160	
BIC	2.493		-0.062		3.751		3.891	
PANEL C: ENNLS with Quadratic Trends								
LGDPPC	1.280	***	1.208	***	1.185	**	0.971	
LGDPPC2	0.311	**	0.160		0.148		0.283	
LGDPPC3			0.065		0.113		0.242	
LGDPPC4					-0.023		-0.406	
LGDPPC5							0.157	
Year	8.591	***	7.822	***	7.750	***	6.966	**
Year Square	-0.002	***	-0.002	***	-0.002	***	-0.002	
Constant	-8480	***	-7720	***	-7650	***	-6890	***
AIC	-19.657		-17.863		-15.877		-14.244	
BIC	-11.628		-7.827		-3.833		-0.192	
PANEL D: CCE								
LGDPPC	0.927	***	1.518	***	0.823	***	0.805	**
LGDPPC2	-0.283	***	-0.832	***	-1.012	***	-0.985	***
LGDPPC3			0.249	***	0.743	***	0.769	**
LGDPPC4					-0.153	***	-0.184	
LGDPPC5							0.007	
Constant	-0.466	***	0.377		-0.210		-0.216	
AIC	-3.474		-7.719		-17.272		-15.282	
BIC	2.656		0.453		-7.057		-3.024	
PANEL E: ENNLS with Demeaned variables								
LGDPPC	0.876	***	1.543	***	1.041	***	0.948	**
LGDPPC2	-0.265	***	-0.886	***	-1.209	***	-1.058	**
LGDPPC3			0.281	**	0.828	***	1.023	**
LGDPPC4					-0.150	***	-0.365	
LGDPPC5							0.045	
Constant	-0.479	***	0.464		0.1163		0.106	
AIC	-8.019		-2.406		-8.304		-6.539	
BIC	-1.997		5.623		1.732		5.506	

(*) null hypothesis can be rejected at 1% significance level.
(**) null hypothesis can be rejected at 5% significance level
(***) null hypothesis can be rejected at 10% significance level

Table 2.5: Baseline Estimations for Western Offshoots

PANEL A: ENNLS								
LGDPPC	2.841	***	17.149	***	-245.893	***	omit	
LGDPPC2	-0.466	***	-5.567	***	135.660	***	-36.801	***
LGDPPC3			0.601	**	-32.659	***	27.167	***
LGDPPC4					2.932	***	-7.427	***
LGDPPC5							0.714	***
Constant	-2.673	***	-15.933	***	167.379	***	27.418	***
AIC	-178		-183		-197		-194	
BIC	-172		-175		-187		-184	
PANEL B: ENNLS with Linear Trends								
LGDPPC	2.953	***	17.286	***	-135.037	**	omit	
LGDPPC2	-0.304	***	-5.411	***	76.022	**	-17.199	
LGDPPC3			0.602	***	-18.655	**	13.349	
LGDPPC4					1.697	**	-3.766	*
LGDPPC5							0.371	*
Year	-0.022	***	-0.022	***	-0.017	***	-0.018	***
Constant	38.890	***	26.275	***	123.021	***	-0.000	***
AIC	-197		-205		-210		-208	
BIC	-189		-195		-198		-196	
PANEL C: ENNLS with Quadratic Trends								
LGDPPC	5.916	***	18.903	***	-119.658	*	omit	
LGDPPC2	-0.861	**	-5.594	***	68.240	*	-13.986	
LGDPPC3			0.572	***	-16.863	**	11.248	
LGDPPC4					1.538	**	-3.241	
LGDPPC5							0.323	
Year	-0.963	***	-0.762		-0.496		-0.557	
Year Square	0.000		0.000		0.000		0.000	
Constant	961.618	*	752.824	***	583.412	***	574.657	***
AIC	-199		-208		-211		-210	
BIC	-191		-198		-199		-198	
PANEL D: CCE								
LGDPPC	2.608	***	3.130	***	3.413	***	3.748	***
LGDPPC2	-0.223	***	-1.086	***	-1.632	**	-1.371	***
LGDPPC3			0.165	***	0.420	**	-0.006	
LGDPPC4					-0.036		0.128	
LGDPPC5							-0.020	
Constant	-0.998	***	0.008	***	0.027		-0.083	
AIC	-125.20		-182.63		-182.78		-181.57	
BIC	-119		-174		-558		-181	
PANEL E: ENNLS with Demeaned variables								
LGDPPC	2.847	***	3.602	***	4.132	***	5.132	***
LGDPPC2	-0.224	***	-1.348	***	-2.275	***	-1.352	
LGDPPC3			0.215	***	0.640	*	-0.756	
LGDPPC4					-0.059		0.467	
LGDPPC5							-0.062	
Constant	-1.373	***	-0.177		-0.175		-0.525	
AIC	-107.39		-118.66		-117.67		-118.88	
BIC	-101		-110		-107		-106	

(*) null hypothesis can be rejected at 1% significance level.
(**) null hypothesis can be rejected at 5% significance level
(***) null hypothesis can be rejected at 10% significance level

Table 2.6: Pairwise Differencing Estimations for China

	Panel A: OLS							
LGDPPC	1.336	***	0.785	**	-0.748	**	1.229	**
LGDPPC2	-0.231	**	-1.022	***	-0.427		-2.015	***
LGDPPC3			0.330	**	-0.518		-1.311	**
LGDPPC4					0.389	**	2.606	***
LGDPPC5							-0.833	**
AIC	-1.979		-10.401		-21.516		-26.884	
BIC	8.236		3.900		-3.128		-4.411	
	PANEL B: ENNLS							
LGDPPC	1.235	***	0.710	**	0.353		0.879	*
LGDPPC2	-0.249	**	-1.114	***	-0.303		-1.549	**
LGDPPC3			0.363	**	-0.405		-1.199	*
LGDPPC4					0.349	**	2.329	**
LGDPPC5							-0.747	*
AIC	-4.912		-11.559		-33.548		-34.500	
BIC	5.065		2.492		-15.481		-12.420	

(*) null hypothesis can be rejected at 1% significance level.
(**) null hypothesis can be rejected at 5% significance level
(***) null hypothesis can be rejected at 10% significance level

Table 2.7: Pairwise Differencing Estimations for Western Offshoots

	Panel A: OLS							
LGDPPC	2.766	***	2.464		-183.149	**	Omitted	
LGDPPC2	-0.401	***	-1.269		-99.064	**	8.803	
LGDPPC3			0.226		23.632	**	-5.599	
LGDPPC4					-2.092	**	1.315	
LGDPPC5							-0.107	
AIC	-227		-233		-236		-240	
BIC	-217		-219		-218		-221	
	PANEL B: ENNLS							
LGDPPC	1.968	***	2.588		178.073	***	Omitted	
LGDPPC2	-0.239	***	-0.836		-93.535	**	41.253	
LGDPPC3			0.116		21.712	**	-28.918	
LGDPPC4					-1.874	**	7.562	
LGDPPC5							-0.698	
AIC	-233		-230		-231		-232	
BIC	-223		-216		-213		-214	

(*) null hypothesis can be rejected at 1% significance level.
(**) null hypothesis can be rejected at 5% significance level
(***) null hypothesis can be rejected at 10% significance level

Table 2.8: Univariate Unit Root Tests for Log - Emission Per Capita

	INDIA	CHINA	COA	WE	WO	AFRICA	LA	CFS	CEE
In levels									
ADF Statistics	-2.163	-3.789	-2.545	-3.854	-5.826	-3.236	-4.401	-2.472	-2.257
Intercept	yes**	yes***	no	yes***	no	yes**	no	yes**	yes**
Trend	yes**	yes***	no	no	no	no	no	no	no
Lag-Length of the Augmented Term	0	1	0	1	1	0	0	2	1
McKinnons' p-value	0.500	0.025	0.000	0.004	0.839	0.002	0.000	0.128	0.190
KPSS Statistics	0.094	0.126*	0.222***	0.222***	0,177**	0.228***	0.235**	0,226***	0,234***
Trend	yes***	yes***	yes***	yes***	yes***	yes***	yes***	yes***	yes***
In First Differences									
ADF Statistics	-8.960				-6.353	-7.028		-1.868	-3.974
Intercept	yes***				no	yes		no	yes**
Trend	no				no	yes***		no	yes**
Lag-Length of the Augmented Term	0				0	0		1	0
McKinnons' p-value	0.000				0.000	0.000		0.059	0.015
KPSS Statistics		0.195	0.132**	0.148**	0.131	0.106	0.054	0.081	0.107
Trend		no	yes***	yes***	no	yes***	yes*	yes***	yes***
Order of Integration by ADF	1	0	0	0	1	0	0	1	1
Order of Integration by KPSS	0	1	2	2	1	1	1	1	1
***	null hypothesis can be rejected at 1% significance level								
**	null hypothesis can be rejected at 5% significance level								
*	null hypothesis can be rejected at 10% significance level								

test, most of the series do not contain a deterministic trend, while for most of them the intercept term is significant. On the other hand the KPSS test finds a significant trend for almost all the series. Lastly, there are parallel findings for the emission and the GDP series for the individual regions in terms of their order of integration which indicates a potential presence of a long term relationship for each cross-sectional unit.

2.A.3. First Generation Panel Unit Root Tests

Univariate unit root tests might suffer from a lack of power due to the small sample size of the dataset. However, since our series for different regions are expected to exhibit some similarities (at least to some degree), both the information contained in the within and between dimension of the panel data set can be exploited to test for unit roots. Our general model is as follows:

$$y_{it} = \rho_i y_{i,t-1} + \alpha_{0i} + \alpha_{1i} t + u_{it}. \tag{2.8}$$

Here, y_{it} is the variable of interest, where y_{i0} is taken as given. The parameters α_{0i}, ρ_i, and α_{1i} are region specific. The error term u_{it} is assumed to be identically and independently distributed (iid) across i and t with zero mean, a finite homoskedastic

Table 2.9: Univariate Unit Root Tests for Log - GDP Per Capita

	INDIA	CHINA	OA	WE	WO	AFRICA	LA	FS	EE
In levels									
ADF Statistics	-3.076	-2.948	-4.502	-5.744	-2.719	-1.329	-1.803	-2.227	-1.395
Intercept	yes***	yes***	yes***	yes***	yes***	yes**	yes***	yes**	no
Trend	no	no	no	no	yes***	no	no	no	no
Lag-Length of the Augmented Term	0	3	0	0	1	1	0	2	1
McKinnons' p-value	1.000	1.000	0.000	0.000	0.233	0.610	0.375	0.200	0.958
KPSS Statistics	0.0.262***	0.262***	0.227***	0.234***	0,132**	0.203**	0.209**	0,192**	0.195**
Trend	yes***	yes***	yes***	yes***	yes***	yes***	yes***	yes***	yes***
In First Differences									
ADF Statistics	-8.826	-5.727			-7.080	-4.254	-6.431	-3.323	-2.960
Intercept	yes	yes			yes***	yes**	yes***	no	yes***
Trend	yes***	yes***			no	no	no	no	no
Lag-Length of the Augmented Term	0	2			0	0	0	0	0
McKinnons' p-value	0.000	0.000			0.000	0.001	0.000	0.001	0.005
KPSS Statistics	0.1314**	0.099	0.0830	0.122**	0.078	0.300	0.306	0.158	0.224
Trend	yes**	yes***	yes***	yes***	no	no	no	no	no
Order of Integration by ADF	1	1	0	0	1	1	1	1	1
Order of Integration by KPSS	2	1	1	2	1	1	1	1	1
***	null hypothesis can be rejected at 1% significance level								
**	null hypothesis can be rejected at 5% significance level								
*	null hypothesis can be rejected at 10% significance level								

variance across t.[18] A noteworthy point is that the common assumption of the first generation unit root tests is cross-sectional independence.

In Table 2.10 the results of the six first generation panel unit root tests for the emission series are presented (GDP per capita will be discussed later). Although there are some conflicting results among the tests, they support the hypothesis that all series have a unit root. Below, these results will be discussed more extensively by focusing on the properties of these tests.

Levin, Lin, and Chu (2002)

Levin et al. (2002) (LLC) test the null that each individual series contains a unit root against the alternative that each individual series is stationary. In terms of equation (1), we test $H_0 : \rho_i = \rho = 0$ against $H_1 : \rho_i = \rho < 0$. Due to the inclusion of individual intercepts and trends we can specify equation (2.8) in an ADF form as follows,

[18] And other regularity conditions, such as a finite fourth order moment.

Table 2.10: p-values from the First Generation Unit Root Tests for the log - Emission Series

Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,0070	0,0000	0,0000	0,0000	0,0000	0,0000
Breitung	0,9590	0,0001	0,8647	0,0000	0,0108	0,0000
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,7582	0,0000	0,0004	0,0000		
ADF – Fisher χ^2	0,6348	0,0000	0,0000	0,0000	0,0000	0,0000
PP – Fischer χ^2	0,8139	0,0000	0,0000	0,0000	0,0000	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,1205	0,0000	0,0000		

Table 2.11: Hypothesis Specification - LLC Test

Specification	Null	Alternative
Model 1a	$H_0(1a): \delta = 0$	$H_1: \delta(1a) < 0$
Model 1b	$H_0(1b): \delta = 0 \; and \; \alpha_{0i} = 0$	$H_1(1b): \delta < 0 \; and \; \alpha_{0i} \in R$
Model 1c	$H_0(1c): \delta = 0 \; and \; \alpha_{1i} = 0$	$H_1(1c): \delta < 0 \; and \; \alpha_{1i} \in R$

$$\begin{aligned}
Model\,1a: \; \Delta y_{it} &= \delta y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \epsilon_{it}, \\
Model\,1b: \; \Delta y_{it} &= \delta y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \alpha_{0i} + \epsilon_{it}, \\
Model\,1c: \; \Delta y_{it} &= \delta y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \alpha_{0i} + \alpha_{1i} t + \epsilon_{it}.
\end{aligned}$$

The parameter δ is equal to $\rho - 1$ and assumed to be constant across the cross sectional units. The term $\sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L}$, containing lagged dependent variables, is included to make the error term asymptotically white noise. For each of these models, the corresponding null and alternative hypotheses are summarized in Table 2.11.

The results of the LLC test for the levels and first differences of log-emission per-capita series of the nine regions are presented in the first row in Table 2.10. For all specifications, whether an individual deterministic trend and/or an intercept is included, the LLC-test indicates stationarity in levels of the emission series.

Asymptotic normality of the LLC-test statistic requires $\sqrt{N/T}$ to go to zero as N

goes to infinity. A large time dimension relative to the cross-section dimension justifies the use of the LLC-test, which is the case in our emission dataset. Indeed, Levin et al. (2002) recommend applying this test to panels with time dimension between 25 and 250 and cross-section dimension between 10 and 250. However, there are some restrictive aspects of the LLC-test. First, the hypothesis formulation is restrictive in the sense that it tests the null hypothesis that all series have a unit root, and rejection indicates that all series are stationary. Secondly, the parameters $\delta = \rho - 1$ are assumed to be homogeneous across regions. To overcome these disadvantages Im et al. (2003) propose a test based on averaging the individual ADF statistics, to be discussed next.

Im, Peseran, and Shin (2003)

Im et al. (2003) (IPS) test the null of all series having a unit root against the alternative that some of the series are stationary, but possibly not all. That is, the null is the same in use of the LLC test; however, under the alternative, the parameter ρ is allowed to be different across units. This is formulated as follows:

$$
\begin{aligned}
H_0 &: \rho_i = 0,\\
H_1 &: \begin{cases} \rho_i < 0 & for\, i = 1, 2, .., N_1, \\ \rho_i = 0 & for\, i = N_1 + 1, \ldots, N. \end{cases}
\end{aligned}
$$

In the ADF form we have three specifications, depending on the inclusion of individual intercepts and deterministic trends:

$$
\begin{aligned}
Model\, 2a &: \Delta y_{it} = \delta_i y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \epsilon_{it},\\
Model\, 2b &: \Delta y_{it} = \delta_i y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \alpha_{0i} + \epsilon_{it},\\
Model\, 2c &: \Delta y_{it} = \delta_i y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \alpha_{0i} + \alpha_{1i} t + \epsilon_{it}.
\end{aligned}
$$

Here, the parameters $\delta_i = \rho_i - 1$ are individual specific as opposed to the LLC test. The corresponding null and alternative hypotheses are summarized in Table 2.12.

The required assumption for consistency of the panel unit root tests is that $\lim_{N \to \infty} N_1/N = a$ where $0 < \alpha \leq 1$. That is, as the number of series grows, the fraction of stationary

Table 2.12: Hypothesis Specification - IPS Test

Specification	Null	Alternative
Model 2a	$H_0(2a)$: $\delta_i = 0$	$H_1(2a)$: $\begin{cases} \delta_i < 0 & for\ i = 1,2,..,N_1 \\ \delta_i = 0 & for\ i = N_1+1,....,N \end{cases}$
Model 2b	$H_0(2b)$: $\delta_i = 0$ and $\alpha_{0i} = 0$	$H_1(2b)$: $\begin{cases} \delta_i < 0 & for\ i = 1,2,..,N_1 \\ \delta_i = 0 & for\ i = N_1+1,....,N \end{cases}$ and $\alpha_{0i} \in R$
Model 2c	$H_0(2c)$: $\delta_i = 0$ and $\alpha_{1i} = 0$	$H_1(2c)$: $\begin{cases} \delta_i < 0 & for\ i = 1,2,..,N_1 \\ \delta_i = 0 & for\ i = N_1+1,....,N \end{cases}$ and $\alpha_{1i} \in R$

series is assumed to stay constant. This is a plausible assumption for cross country studies. Results of this test are also presented in Table 2.10. If we assume individual deterministic trends, opposite to the LLC-test, the results of the IPS-test (Table 2.10, row 3) implies that all series have a unit root. If a trend is excluded, the IPS-test implies stationarity in levels of the series. It seems that including trends changes the results dramatically. Indeed, this point is highlighted by Breitung (2000), as discussed in the following section.

Breitung (2000)

Both the IPS and the LLC tests have the disadvantage of requiring T to be large relative to N. Besides, they are sensitive to the specification of the deterministic trend being individual specific or not. Breitung (2000) argues that the IPS and the LLC tests have size distortions as N/T increases; that is, they reject the null hypothesis too often. Furthermore, there is a substantial loss of power if individual deterministic trends are included. The panel unit root test proposed by Breitung (2000) is free of these criticisms. The hypothesis formulation is the same with the LLC test and it assumes a common unit root process among the series, similar as the LLC test.

Results are presented in the second row of Table 2.10. When the individual deterministic trends and intercepts are excluded, the results of the Breitung-test do not conflict with the previous results. If individual intercepts are included, the Breitung-test does not reject the null hypothesis of a unit root, while the LLC and IPS tests do reject. If individual deterministic trends are included, then the test indicates the presence of a unit root for each series. Actually, if Breitung's criticism about the IPS and LLC tests, namely, that they suffer from loss of power due to the inclusion of individual intercepts and trends, is the reason why there are conflicting results, we would expect the former tests not to reject the null of a unit root, while the Breitung test rejects. However, the situation is just the other way around. The problem could be a size distortion caused by large N compared to T; however, in our case T seems to be large compared to N. So, the conflicting results might not be based on the arguments against the LLC and IPS tests, as presented by Breitung (2000).

ADF and Philips-Perron (PP) Fisher Chi-square test

The ADF Fisher Chi-square test assumes individual unit root processes under the null of a unit root. It is similar to the IPS test in terms of its hypothesis formulation and incorporating the idea to combine the information from individual unit root tests. Its advantage over the IPS is that it allows for unbalanced panels and different lag lengths in individual ADF regressions. For the model with individual trends, results support the hypothesis that all series have a unit root (Table 2.10 – row 4); however, if trends are excluded, it implies stationarity for the levels. The PP Fisher Chi-square test proposed by Choi (2001) is suggested when N is large. This test also supports the results of the Fisher Chi-Square test (Table 2.10 – row 5).

Hadri test

Hadri (2000) proposed a Lagrange multiplier test where the null hypothesis is the stationarity of all series against the alternative of a unit root in the panel. It is based on the KPSS univariate unit root test. Results are presented in the last row of Table 2.10. For all specifications, the Hadri test suggests that levels and first differences are non-stationary. As a result, the conclusion of the Hadri-test about the first differences conflicts with the previous tests. The reason is most likely related to the alternative hypothesis of all series having a unit root. Even if one series is non-stationary, the Hadri-test rejects the null that all series are stationary. This result may not be surprising; indeed, the univariate analysis indicates that some series could be non-stationary.

Table 2.13: p-values from the First Generation Unit Root Tests for the log - GDP Series

Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,1715	0,0000	0,0000	0,0000	1,0000	0,0000
Breitung	0,4470	0,0000	0,3033	0,0000	0,7748	0,0000
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,9965	0,0000	0,8101	0,0000		
ADF – Fisher χ^2	0,9695	0,0000	0,0004	0,0000	1,0000	0,0000
ADF – Fisher χ^2	0,9971	0,0000	0,0084	0,0000	1,0000	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0000	0,0000	0,0001		

Table 2.14: p-values from the Cross-sectional Independence Tests

Series	Breusch and Pagan (1980)	Pesaran (2004)	Friedman (1937)	Frees (1995, 2004)
LN(Emission PC)	0.000	0.000	0.000	0.000
LN(GDP PC)	0.000	0.000	0.000	0.000

Log GDP per capita

In Table 2.13, the first generation analysis is replicated for the logarithm of GDP per-capita series. Results are very similar except for a few cases. Like for the emission series, the results are very sensitive to the type of the test conducted and the specification of deterministic terms. Again, the univariate analysis indicates potential presence of deterministic trends in most of our series. In that sense, model 1c seems to be more reliable. For this specification, there is very few conflicting results across different tests, both for the emission and GDP series, which are all found out to be I(1).

2.A.4. Cross-sectional Dependence

For cross-sectional dependence, two specifications are considered, depending on the assumed dependence. The first one ignores common factors by assuming dependence only through the cross correlations in the errors. We conduct the tests described in De Hoyos and Sarafidis (2007) to investigate the presence of cross-sectional dependence in our series. The null hypothesis for all tests is cross-sectional independence. In Table 2.14, results are presented in which all tests strongly reject the null, and hence indicate that the error terms, ε_{it} in equation (2.8), are correlated across cross-sectional units.[19]

The second specification incorporating cross-sectional dependence is a factor model,

[19] The references in this table are the references used by De Hoyos and Sarafidis (2007). We refer to this paper for these references.

where the cross sectional dependence is due to some unobserved common factors. Identifying the factors is problematic for our data set, due to having nine series for each variable and a time dimension of 59. A desirable sample size/variable ratio is at least ten. Another sample size problem arises when selecting the number of factors to be extracted. For example, the selection criteria proposed in Bai and Ng (2002) are inclined to choose far more factors than the data generating process assumes in their Monte-Carlo simulations. However, these problems do not make it impossible to get some valuable insights by a factor analysis, although one should be cautious in interpreting the results. The following analysis is mainly in line with the discussions about principle component analysis in Anderson et al. (2006).

We start with checking whether our panel variables satisfy some conditions for a factor analysis. Firstly, as discussed in Anderson et al. (2006), the minimum required sample size/variable ratio is five which is satisfied in our case. However, it is also mentioned that the desirable ratio is at least ten, which is not satisfied in our case. Secondly, a substantial number of correlations should be higher that 0.30. This condition is satisfied as it can be seen from the correlation matrix presented for both the emission and GDP series in Table 2.15. Secondly, we apply the Bartlett Sphericity test to see whether there are equal correlations which invalidate a latent structure. As presented in Table 2.16, both for the GDP and the emission series, equality of all correlations is rejected. Lastly, we check the Kaiser-Meyer-Olkin measure of sampling adequacy which should be larger than 0.5 for each variable. In Table 2.17, we show that both variables satisfy this condition. Therefore, it seems appropriate to apply a factor analysis according to these conventional methods.

Next, in order to determine the number of common factors, we use parallel analysis and scree-plots. When the principle component analysis is performed with the levels of variables, for both the emission and the GDP series, the parallel analysis slightly favors one common factor, while the scree-plot slightly favors two common factors as can be seen in Figures 2.12 and 2.13. On the other hand, when the analysis is performed in first differences, the parallel analysis slightly favors two common factor, while the scree-plot slightly favors three common factors for each variable, as can be seen in Figures 2.14 and 2.15. As a result, we will continue our analysis for each case, assuming one, two, or three common factors.

Table 2.15: Correlation Matrix

Emission	IND	CHI	OA	WO	AFR	LA	FU	EE	WE
IND	1.000								
CHI	0.938	1.000							
OA	0.913	0.914	1.000						
WO	0.999	0.934	0.912	1.000					
AFR	0.876	0.900	0.973	0.879	1.000				
LA	0.930	0.939	0.964	0.933	0.970	1.000			
FU	0.479	0.627	0.731	0.479	0.804	0.704	1.000		
EE	0.453	0.603	0.728	0.451	0.800	0.692	0.978	1.000	
WE	0.587	0.704	0.855	0.581	0.848	0.772	0.878	0.906	1.000

GDP	IND	CHI	OA	WO	AFR	LA	FU	EE	WE
IND	1.000								
CHI	0.991	1.000							
OA	0.887	0.888	1.000						
WO	0.996	0.993	0.914	1.000					
AFR	0.748	0.743	0.944	0.780	1.000				
LA	0.857	0.861	0.991	0.888	0.961	1.000			
FU	0.369	0.375	0.676	0.412	0.841	0.703	1.000		
EE	0.770	0.771	0.946	0.805	0.990	0.960	0.827	1.000	
WE	0.909	0.908	0.997	0.932	0.936	0.985	0.661	0.945	1.000

Table 2.16: Bartlett Sphericity Test

Null Hypothesis: All correlations are equal		
	Emission	GDP
Lawley Chi-square	973	1066
p-value	0.000	0.000

Table 2.17: Kayser-Meyer-Olkin Sampling Adequacy

	Emission	GDP
IND	0.722	0.729
CHI	0.939	0.932
OA	0.856	0.786
WO	0.719	0.755
AFR	0.793	0.782
LA	0.933	0.891
FU	0.879	0.766
EE	0.758	0.876
WE	0.854	0.805

Figure 2.12: Number of Factors in Level of Emission Series

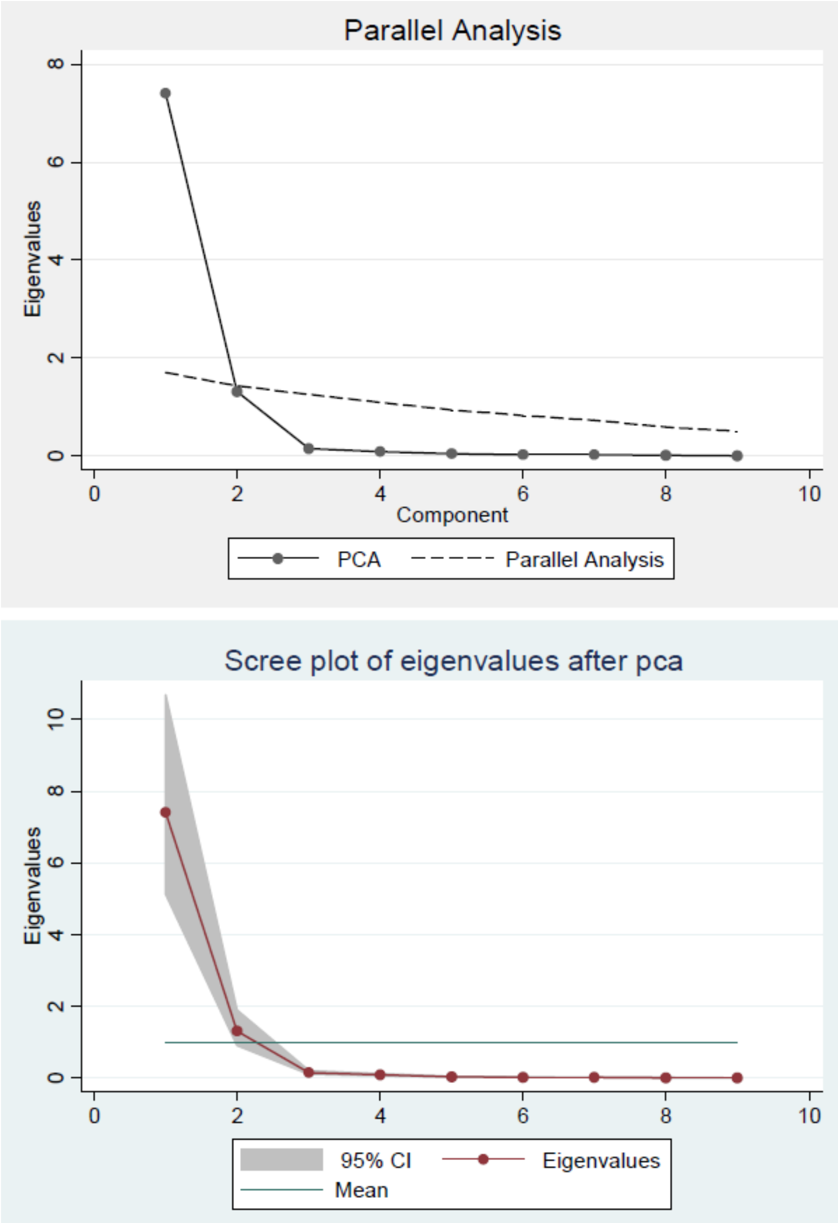
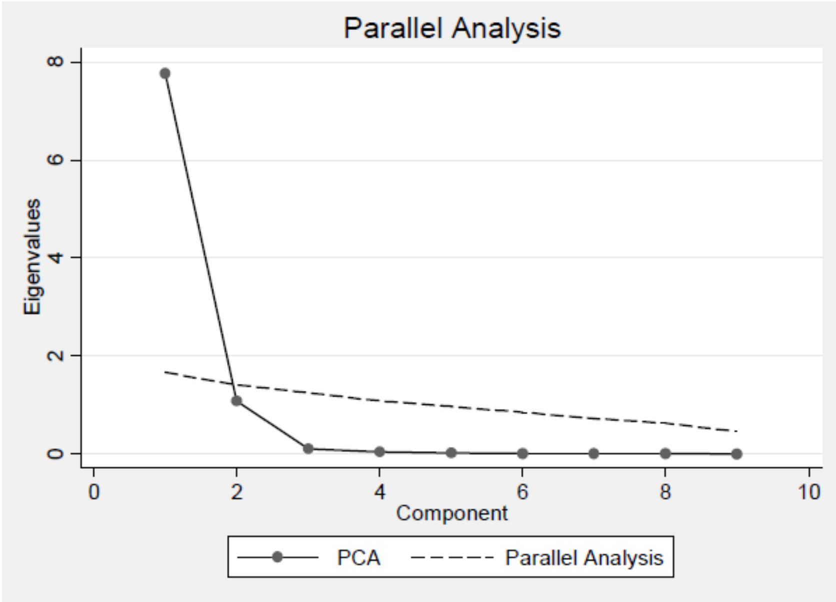


Figure 2.13: Number of Factors in Level of GDP Series



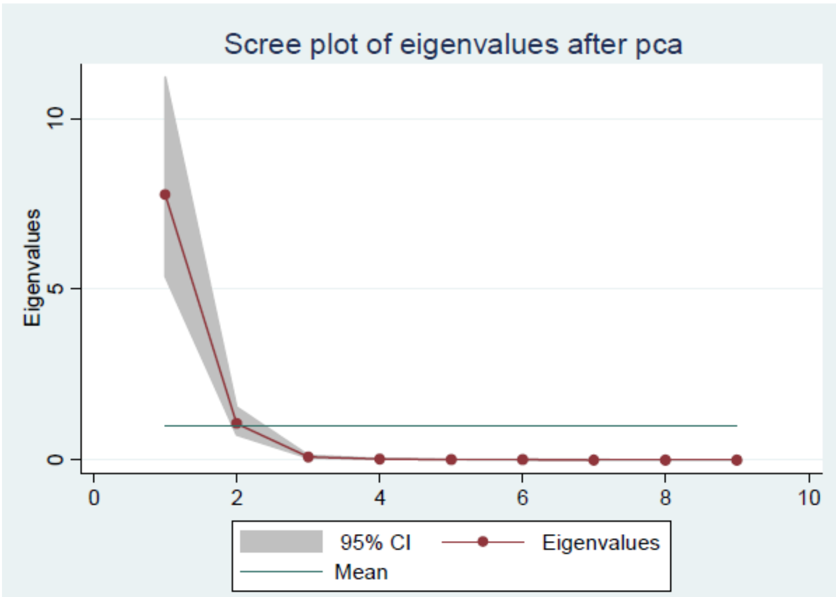


Figure 2.14: Number of Factors in First Differences of Emission Series

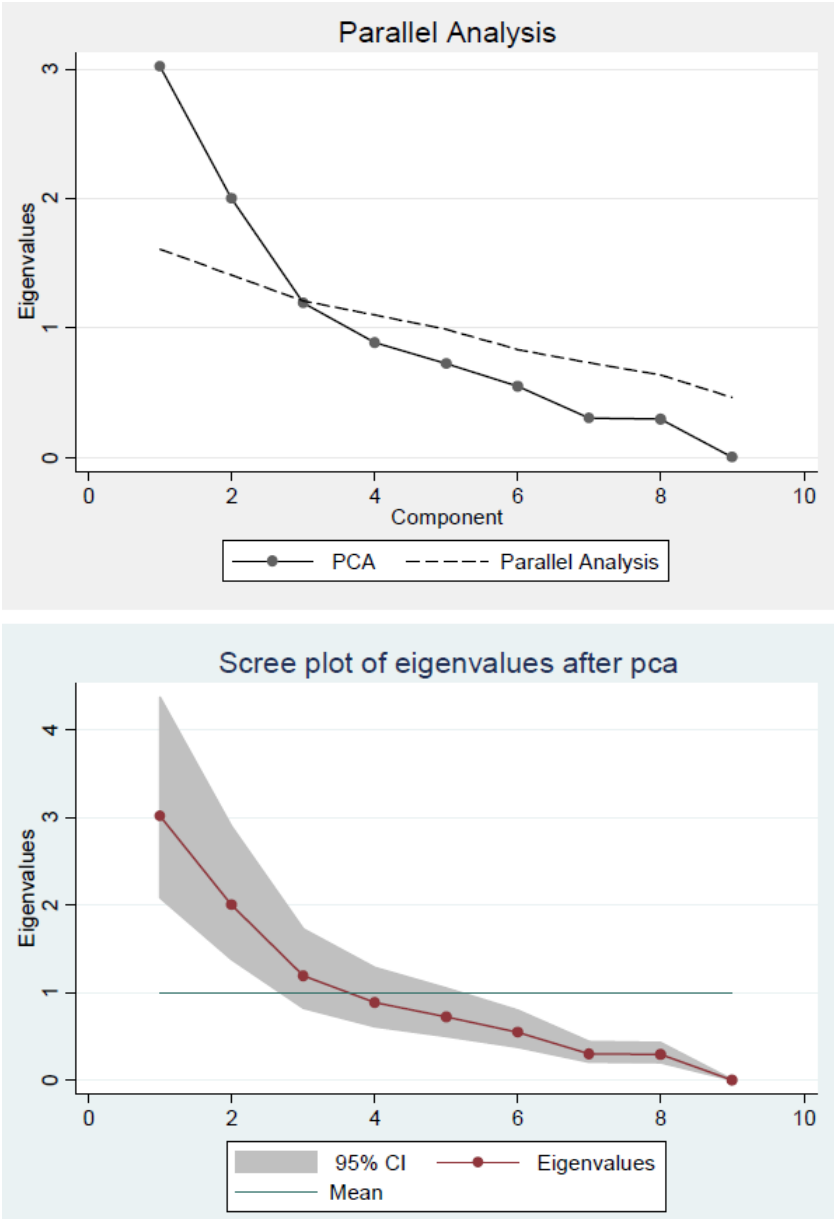
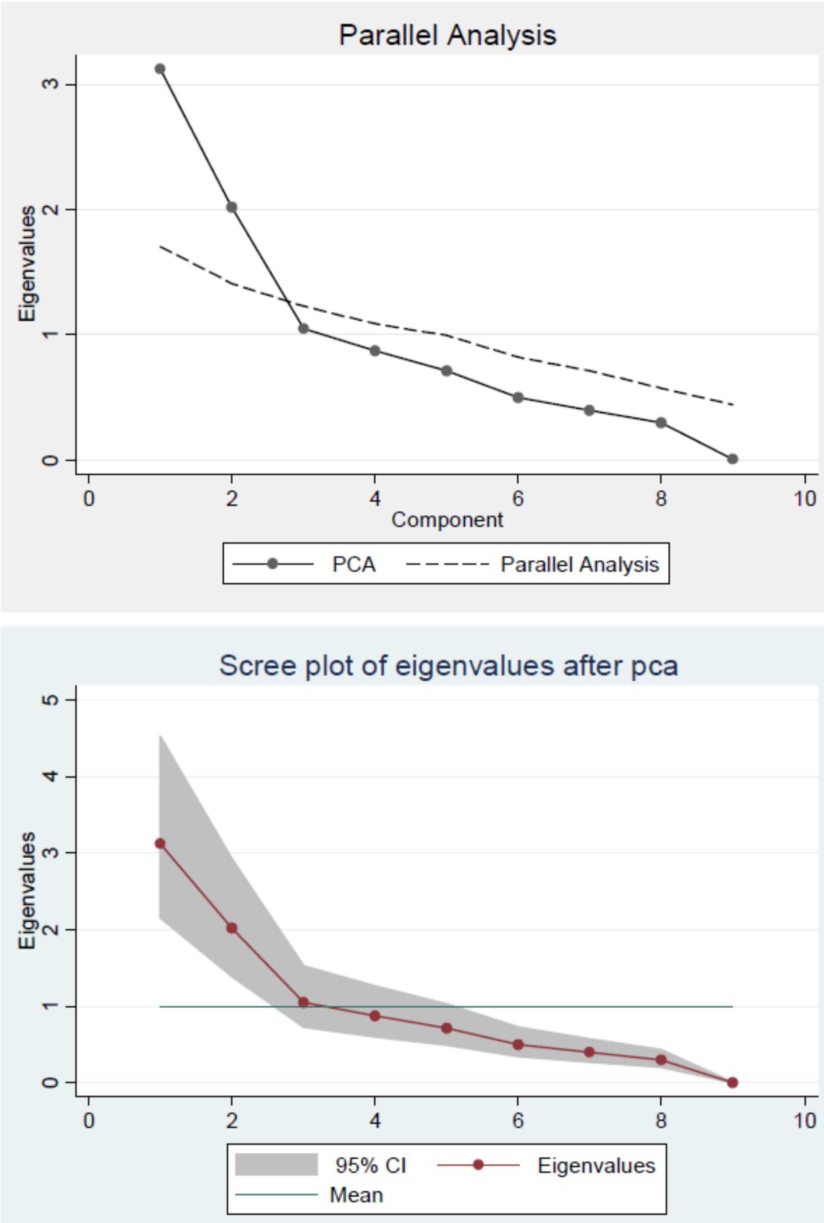


Figure 2.15: Number of Factors in First Differences of GDP Series



We also present for both variables the results of the first generation unit root tests, assuming that the cross-sectional dependence in our variables is totally due to the factors common to the cross-sectional units. The test results are presented in Tables 2.18 to 2.21. The analysis shows that both the common factors and the idosyncratic components are I(1) processes for both the emission series and the GDP series. This result is robust whether assuming one, two, or three common factors, extracting the factors by first differencing or by using the levels of the variables.

2.A.5. Second Generation Panel Unit Root Tests

As presented in the previous section, there is a strong evidence of cross-sectional dependence in our panel variables. Therefore, in this section, we apply second generation unit root tests allowing for cross-sectional dependence. Firstly, assuming one common factor for each variable, we apply the cross sectional augmented IPS (CIPS) test suggested by Pesaran (2007).[20] One advantage of this test is that it does not require to estimate the factors. Secondly, we apply the tests suggested by Moon and Perron (2004), and Bai and Ng (2004) which allow for more than one common factor, and therefore require the factor loadings to be estimated. The results of the latter one are presented in the main text.

Pesaran (2007)

In case of one common factor, Pesaran (2007) suggests to proxy the common factor by the cross-section averages $\bar{y}_t$ and lags. Without the serial correlation, $\bar{y}_t$ and $\bar{y}_{t-1}$ are sufficient to proxy the common factor. Therefore, by assuming one common factor, the model given by equation (2.8) can be expressed in ADF form as follows:

$$\Delta y_{it} = \delta_i y_{i,t-1} + \alpha_{0i} + \alpha_{1i} t + b_i \bar{y}_{t-1} + c_i \bar{y}_t + \xi_{it}, \tag{2.9}$$

where the error term ε_{it} in (2.8) is modeled as $\varepsilon_{it} = b_i \bar{y}_{t-1} + c_i \bar{y}_t + \xi_{it}$. In order to deal with serial correlation, these regressions can be augmented with the lags of Δy_{it} and $\Delta \bar{y}_t$. The null hypothesis, presence of a unit root such that $\delta_i = 0$ for all i, is tested against the stationary alternative that $\delta_i < 0$ for a fixed fraction of the cross-sectional units.

The test statistic proposed by Pesaran (2007) is the cross-sectional augmented version of the IPS-test. The IPS-test statistic is the average of the univariate ADF tests applied

[20] IPS stands for the Im-Peseran-Shin test. See Im et al. (2003).

Table 2.18: p-values from the First Generation Unit Root Tests for factors of Log Emission-pc Series

For one common factor identified with levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
ADF	0.7628	0.0005	0,0296	0.0001	0.2559	0.001
KPSS	0,2376***	0,1513**	0,8139***	0,7918***		
For two common factors identified with levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,1088	0,0037	0,0000		0.1331	0,0000
Breitung	0,8982	0,0630				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,8141	0,0001	0.0037			
ADF - Fisher χ^2	0,8858	0,0002	0,0041		0.1145	0,0000
PP - Fisher χ^2	0,8502	0.0000	0,0009		0.0028	
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0.000	0.0843	0.000	0.0006		
For two common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,0141	0,0000	0,7357	0,0000	0.2829	0,0000
Breitung	1.000	0,0053				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,4160	0,0000	0,6116	0,0000		
ADF - Fisher χ^2	0,4930	0,0000	0,5183	0,0000	0.0934	0,0000
PP - Fisher χ^2	0,5607	0,0000	0,4041	0,0000	0.0642	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0674	0,1415	0,0000		
For three common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,0640	0,0000	0,7596	0,0000	0.2013	0,0000
Breitung	0,9994	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,6978	0,0000	0,7285	0,0000		
ADF - Fisher χ^2	0,8012	0,0000	0,7048	0,0000	0.1495	0,0000
PP - Fisher χ^2	0,7965	0,0000	0,5652	0,0000	0.1009	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,1993	0,1077			

Table 2.19: p-values from the First Generation Unit Root Tests for Idiosyncratic components of the Log Emission-pc Series

For one common factor identified with levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,0000		0,0000		0.000	
Breitung	1.0000	0,0125				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,0138	0,0000	0.0000			
ADF - Fisher χ^2	0.0735	0,0000	0,0000		0.0000	
PP - Fisher χ^2	0,3377	0.0000	0,0000		0.0000	
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0.000	0.0000	0.000	0.0000		
For two common factors identified with the levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,1715	0,0001	0,0000	0,0000	0,0000	
Breitung	0,4470	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,8580	0,0000	0,9901	0,0000		
ADF - Fisher χ^2	0,8860	0,0000	0,0000	0,0000	0,0000	
PP - Fisher χ^2	0,9426	0,0000	0,0000	0,0000	0,0000	
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0029	0,0000	0,0000		
For two common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,0027		0,9999	0,0000	0.3265	0,0000
Breitung	1.0000	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,9973	0,0000	1.0000	0,0000		
ADF - Fisher χ^2	0.9999	0,0000	1.0000	0,0000	0.9667	0,0000
PP - Fisher χ^2	1.0000	0,0000	1.0000	0,0000	0.5891	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0008	0,0102	0,0000		
For three common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,2213	0,0000	0,9998	0,0000	0.3223	0,0000
Breitung	1.0000	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	1.0000	0,0000	1,0000	0,0000		
ADF - Fisher χ^2	1.0000	0,0000	1.0000	0,0000	0.8833	0,0000
PP - Fisher χ^2	1.0000	0,0000	0,9998	0,0000	0.4976	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,3993	0,0334	0,0000		

Table 2.20: p-values from the First Generation Unit Root Tests for factors of Log GDP-pc Series

For one common factor identified with levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
ADF	0.7482	0.0595	0,9082	0.0111	0.7193	0.0992
KPSS	0,1943**	0,1677**	0,9064			
For two common factors identified with levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,4322	0,0617	0,9739	0.0267	0.5216	0,0000
Breitung	0,9991	0,0037				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,9363	0,0061	0.9840	0.0018		
ADF - Fisher χ^2	0,9640	0,0084	0,3754	0.0027	0.0818	0,0000
PP - Fisher χ^2	0,9474	0.0062	0,6253	0.0024	0.1429	0.0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0.000	0.0436	0.0001	0.0630		
For two common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,6689	0,0000	0,8361	0,0000	0.2769	0,0000
Breitung	1.0000	0,0029				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,9993	0,0000	0,9231	0,0000		
ADF - Fisher χ^2	0,9380	0,0000	0,9543	0,0000	0.6327	0,0000
PP - Fisher χ^2	0,9656	0,0000	0,8685	0,0000	0.5003	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0038	0,0435	0,0002		
For three common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,7665	0,0000	0,8412	0,0000	0.1968	0,0000
Breitung	0,9999	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,9991	0,0000	0,9226	0,0000		
ADF - Fisher χ^2	0,9961	0,0000	0,9694	0,0000	0.6445	0,0000
PP - Fisher χ^2	0,9937	0,0000	0,8945	0,0000	0.4774	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,1993	0,1077			

Table 2.21: p-values from the First Generation Unit Root Tests for idiosyncratic components of Log GDP-pc Series

For one common factor identified with levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,6924	0.0000	0,0006		0.000	
Breitung	1.0000	0,0008				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,9976	0,0000	0.8852	0.0000		
ADF - Fisher χ^2	1.0000	0,0001	0,9905	0.0000	0.0000	
PP - Fisher χ^2	0,9984	0.0003	0,9995	0.0000	0.0000	
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0.000	0.0000	0.000	0.0000		
For two common factors identified with the levels of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,5949	0,0000	1.0000	0,0000	0,0000	
Breitung	1.0000	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	1.0000	0,0000	1.0000	0,0000		
ADF - Fisher χ^2	1.0000	0,0000	1.0000	0,0000	0,0000	
PP - Fisher χ^2	1.0000	0,0000	1.0000	0,0000	0,0000	
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0029	0,0000	0,0000		
For two common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,2514		0,9868	0,0000	0.1689	0,0000
Breitung	1.0000	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,9998	0,0000	0.9996	0,0000		
ADF - Fisher χ^2	0.9999	0,0000	1.0000	0,0000	0.9040	0,0000
PP - Fisher χ^2	1.0000	0,0000	0.9974	0,0000	0.6201	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0008	0,0003	0,0000		
For three common factors identified with first differences of the variables						
Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,2379	0,0000	0,9571	0,0000	0.0584	0,0000
Breitung	1.0000	0,0000				
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0.9455	0,0000	0.9828	0,0000		
ADF - Fisher χ^2	0.9950	0,0000	0.9957	0,0000	0.5349	0,0000
PP - Fisher χ^2	0.9801	0,0000	0,9407	0,0000	0.2226	0,0000
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,3993	0,0334	0,1039		

Table 2.22: Pesaran (2007) - p values from the CIPS Test

Series	Constant & Trend	Lag Number	Constant	Lag Number
Ln Emission-pc	0,545	1	0,692	1
ΔLn Emission-pc	0,000	1	0,000	1
Ln GDP-pc	0,887	1	0,997	1
ΔLn GDP-pc	0,000	1	0,000	1

to each individual series. Therefore, the only required modification for Pesaran's test is a slight modification of the standard error of the unit root coefficient estimates due to the presence of the cross-sectional dependence. Pesaran (2007) shows that the asymptotic distribution of the Cross sectional augmented ADF (CADF) statistic proposed is free of nuisance parameters introduced with the factor structure, as $N \to \infty$, followed by $T \to \infty$, or as $N \to \infty$ and fixed $T > 3$. The Monte-Carlo simulation performed by Pesaran (2007) shows that the test has satisfactorily small sample properties even for very small sample sizes. More specifically, the size properties are satisfactory even for $N = T = 10$, and the power of the test increases with N, when $N > 30$.

Results of the tests proposed in Pesaran (2007), which are presented in Table 2.22, indicate that all series are $I(1)$ processes. Results are robust to including a deterministic trend.

Moon and Perron (2004)

The approach followed by Moon and Perron (2004) is similar to Pesaran (2007), in assuming that the error terms follow a factor structure. In case of one common factor, the data generating process is identical with the one in Pesaran (2007). Also the hypothesis formulation is the same as Pesaran (2007). However, Moon and Perron (2004) allow for more than one common factor. Therefore, the data generating process is given by

$$\Delta y_{it} = \delta_i y_{i,t-1} + \alpha_{0i} + \alpha_{1i} t + \gamma_i' f_t + \xi_{it}, \tag{2.10}$$

for some vector of factors f_t, with corresponding vector of parameters γ_i. In contrast to Pesaran (2007), the factor loadings have to be estimated. Moon and Perron (2004) use the information criteria suggested by Bai and Ng (2004) in order to determine the number of common factors.

In Table 2.23, we present the results of the Moon and Perron (2004) test in case of one, two, and three common factors. For all cases, the test finds a unit root if trends

Table 2.23: Moon and Perron (2004) Second Generation Unit Root Test

	Number of Factors = 1			
	without trend		with trend	
	ta statistic	tb statistic	ta statistic	tb statistic
ln(gdp-pc)	0	0.002	0.9988	0.9986
ln(emm-pc)	0	0	0.139	0.074
Δln(gdp-pc)	0	0	0	0
Δln(emm-pc)	0	0	0	0
	Number of Factors = 2			
	without trend		with trend	
	ta statistic	tb statistic	ta statistic	tb statistic
ln(gdp-pc)	0	0	0.22	0.218
ln(emm-pc)	0	0	0.28	0.298
Δln(gdp-pc)	0	0	0	0
Δln(emm-pc)	0	0	0	0
	Number of Factors = 3			
	without trend		with trend	
	ta statistic	tb statistic	ta statistic	tb statistic
ln(gdp-pc)	0	0	0.241	0.255
ln(emm-pc)	0	0	0.211	0.236
Δln(gdp-pc)	0	0	0	0
Δln(emm-pc)	0	0	0	0

Table 2.24: Moon, Perron & Philips (2007) Point Optimal Unit Root Test Results

Series	Constant	Constan & Trend
LN(Emission PC)	49.634	8.240
Δ LN(Emission PC)	-5.595	-3.449
LN(GDP PC)	70.645	7.573
Δ LN(GDP PC)	-5.757	-3.352

are included; however, for the case where the trends are excluded, stationarity is the result. This is valid for both the GDP and the emission data. As noticed in Table 2.23, the results differ substantially depending on the inclusion of deterministic trends. Moon and Perron (2004) check the asymptotic power of the proposed test against some local alternatives with a near unit root hypothesis. It is shown that when individual deterministic trends are included, the proposed test does not possess any asymptotic power against local alternatives. This issue is further investigated in Moon et al. (2007), who use the same procedure in order to compare different tests proposed in the literature. Their finding is that for almost all tests, incidental trends cause a loss of power. In the same paper, the authors proposed a point optimal test which achieves the power envelop even in the case of incidental trends. The results for this test are presented in Table 2.24. Test results imply first difference stationarity for all series.

Bai and Ng (2004)

One problem with the panel unit root tests that may cause a substantial bias is the presence of cross-cointegration across panel units which should be distinguished from the case where the errors are cross correlated. Therefore, the underlying null hypothesis in Moon and Perron (2004) and Pesaran (2007) can be restated as all series have a unit root and there is no cointegrating relationship among all series. Bai and Ng (2004) suggest a multi-factor framework called “Panel Analysis of Non-stationarity in Idiosyncratic and Common Components” (PANIC) where the factors and idiosyncratic components are analyzed separately and hence allows for cross-unit cointegration. Furthermore, this method allows testing for a unit root in the factors. The results of this test are presented in the main text.

2.A.6. Pairwise Differencing and Unit Roots

Taking into acount that the pairwise differencing eliminates the cross-sectional dependence (under the assumption of common time effects), we apply first generation unit root tests for the pairwise differences of the emission and GDP series which are presented in Table 2.25 and 2.26. Results are in line with the factor analysis that our pairwise differenced variables are all I(1).

Table 2.25: First Generation Unit Root Tests for Pairwise Differenced Emission Series

Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,0022	0,0000	0,0203	0,0000	0,0000	0,0000
Breitung	0,9998	0,0001	0,9997	0,0000	0,0000	0,0000
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,0041	0,0000	0,1044	0,0000		
ADF – Fisher χ^2	0,6598	0,0000	0,1503	0,0000		
PP – Fisher χ^2	0,0009	0,0000	0,0000	0,0000		
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0000	0,0000	0,0000		

Table 2.26: First Generation Unit Root Tests Pairwise Differenced GDP Series

Specification	Ind Trend & Intercept		Individual Intercept		None	
Series	Level	First Dif.	Level	First Dif.	Level	First Dif.
Hypothesis	H_0 (1c) against H_1 (1c)		H_0 (1b) against H_1 (1b)		H_0 (1a) against H_1 (1a)	
LLC	0,0059	0,0000	0,0073	0,0000	0,0349	0,0000
Breitung	1.0000	0,0001	0,9955	0,0000	0,0061	0,0000
Hypothesis	H_0 (2c) against H_1 (2c)		H_0 (2b) against H_1 (2b)		H_0 (2a) against H_1 (2a)	
IPS	0,5587	0,0000	0,9172	0,0000		
ADF – Fisher χ^2	0,0160	0,0000	0,3431	0,1650		
PP – Fisher χ^2	0,9185	0,0000	0,0001	0,0000		
Hypothesis	Stationarity around a trend against all series have a unit root		Stationarity around a level against all series have a unit root			
Hadri	0,0000	0,0000	0,0000	0,0000		

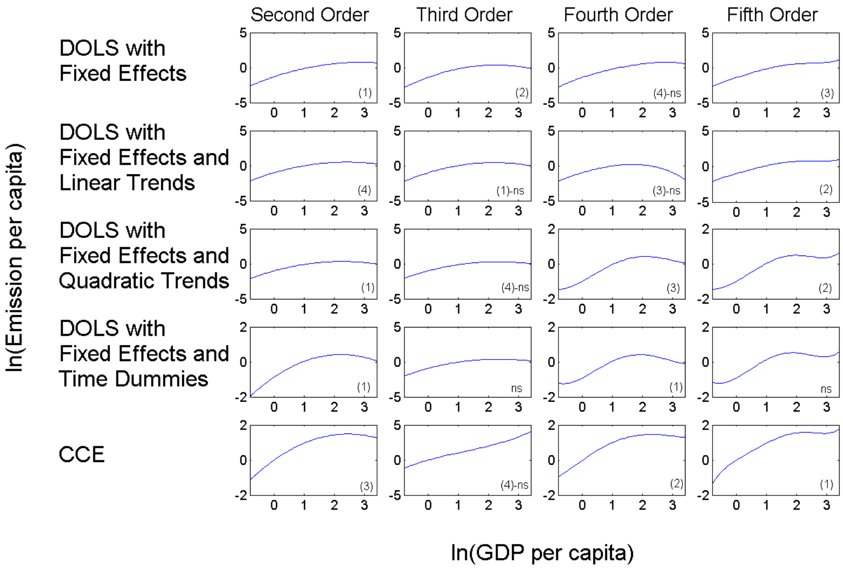
2.B. Estimation Results under Homogeneity

In the main text, we only consider the heterogeneous case. In this section, we consider the homogeneous case, i.e.,

$$y_{it} = f(x_{it}) + \lambda(t) + \varepsilon_{it}, \tag{2.11}$$

with f and λ not region-specific. Our baseline estimations under the assumption of homogeneous income effects include DOLS estimations, differing in their specifications of the common deterministic trends, and the CCE estimations which control for the common stochastic trends. In Figure 2.16, the within sample plots of the baseline estimations for different polynomials of the hypothesized emission-income relation are depicted. The number "(#)" on the bottom-right corner of each graph indicates the rank of preference by BIC, and the curves for which the highest order term of the polynomial is not significant is indicated with "ns". The underlying estimation results are reported in Table 2.27. The predicted curves are in line with the EKC hypothesis for all estimations in terms of the income effect, since the curves shows the income effects. While the best fit functional form of all the DOLS estimations are the quadratic equations, selected on

Figure 2.16: Baseline Estimations under Homogeneity



the basis of the BIC, for the CCE estimation, it is the fourth order polynomial. This predicts a stabilization in the emissions, but no clear turning point. This is also the case for the DOLS estimations with the fourth order polynomial. On the other hand, these results are counterintuitive when interpreting the income effect as a positive scale effect.

Applying the pairwise differencing strategy does not change the results. Figure 2.17, and the underlying estimation results in Table 2.28, show that a fixed effect estimation indicates an inverted U-shaped income-emission relation for all polynomial specifications, except the fifth order polynomial, which is also the best fit curve according to the BIC. However, when the fixed effects estimation is performed in a DOLS context to increase efficiency, the fifth order term is not significant. Therefore, the pairwise differencing estimations support the baseline estimations. To sum up, the estimation strategies adopted under the assumption of homogeneity strongly support the EKC hypothesis in terms of income, but are counterintuitive when interpreting income as a scale effect.

In order to analyze the decomposition of the total emissions into an income and a time effect, we use the estimation results of the DOLS pairwise differencing estimation with fixed effects for China and Western Offshoots, presented in Figure 2.18. The pairwise differencing estimation does not identify the levels of these curves. Therefore, we normalize the curves such that the level of the sample average for each curve is equal to the average level of the observed emission in that region. The slope of the curves show

Table 2.27: : Baseline Estimations under Homogeneity

Panel A: DOLS with Fixed Effects								
LGDPPC	1.432	***	1.566	***	1.429	***	1.398	***
LGDPPC2	-0.251	***	-0.355	***	-0.330	***	-0.135	
LGDPPC3			0.013		0.085		0.050	
LGDPPC4					-0.019	**	-0.079	
LGDPPC5							0.017	**
Constant	-1.299	***	-1.294	***	-1.387	***	-1.427	***
Lags	4		4		1		1	
Leads	4		4		1		2	
BIC	-492		-490		-391		-444	
PANEL B: DOLS with Fixed Effects and Linear Time Trends								
LGDPPC	1.254	***	1.310	***	1.288	***	1.226	***
LGDPPC2	-0.261	***	-0.288	***	-0.282	**	-0.109	***
LGDPPC3			-0.004		-0.026		0.054	
LGDPPC4					-0.008		-0.080	
LGDPPC5							0.017	***
Year	0.006	***	0.006	***	0.006	***	0.004	***
Leads	4		4		4		2	
Lags	4		4		2		1	
BIC	-439		-398		-359		-371	
PANEL C: DOLS with Fixed Effects and Quadratic Time Trends								
LGDPPC	1.183	***	1.137	***	0.941	***	0.942	***
LGDPPC2	-0.264	***	-0.254	***	0.237	***	0.311	***
LGDPPC3			0.002		-0.253	***	-0.024	*
LGDPPC4					0.038	***	-0.024	
LGDPPC5							0.014	**
Year	1.700	***	1.726	***	1.930	***	1.860	***
Year Square	-0.000	***	-0.000	***	-0.001	***	-0.001	***
Leads	4		4		2		2	
Lags	2		2		1		1	
BIC	-577		-548		-553		-565	
PANEL C: DOLS with Fixed Effects and Time Dummies								
LGDPPC	1.161	***	1.159	***	0.8893	***	0.925	***
LGDPPC2	-0.263	***	-0.274	***	0.353	***	0.457	***
LGDPPC3			0.007		-0.338	***	-0.375	***
LGDPPC4					0.052	***	0.031	
LGDPPC5							0.008	
Constant	-0.864	***	-0.850	***	-0.916	***	-1.506	***
Leads	4		4		4		3	
Lags	3		2		2		1	
BIC								
PANEL C: CCE								
LGDPPC	1.215	***	1.213	***	1.200	***	1.089	***
LGDPPC2	-0.249	***	-0.246	***	-0.124	**	-0.336	***
LGDPPC3			-0.001		-0.093	**	-0.340	***
LGDPPC4					0.017	**	-0.196	***
LGDPPC5							0.032	***
Constant	-0.000	***	-0.000	***	-0.000	***	-0.000	***
BIC	-560		-556		-558		-593	

(*) null hypothesis can be rejected at 1% significance level.
(**) null hypothesis can be rejected at 5% significance level
(***) null hypothesis can be rejected at 10% significance level

Figure 2.17: Pairwise Differencing Estimations under Homogeneity

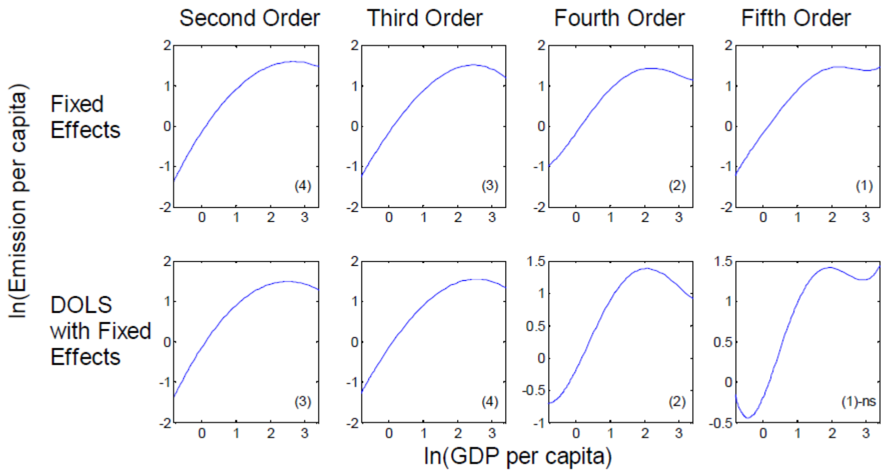


Table 2.28: Pariwise Differencing Estimations Homogeneity

Panel A: Fixed Effects								
LGDPPC	1.299	***	1.226	***	1.190	***	1.164	***
LGDPPC2	-0.242	***	-0.168	***	0.052		-0.071	
LGDPPC3			-0.022	***	-0.180	***	0.030	
LGDPPC4					0.028	***	-0.071	**
LGDPPC5							0.015	***
Constant	-0.151		-0.165	***	-0.182	***	-0.185	***
AIC	-424		-439		-469		-479	
BIC	-411		-422		-448		-453	
PANEL B: DOLS with Fixed Effects								
LGDPPC	1.304	***	1.245	***	1.053	***	0.994	**
LGDPPC2	-0.260	***	-0.194	***	0.266	***	0.683	***
LGDPPC3			-0.013	**	-0.284	***	-0.635	***
LGDPPC4					0.042	***	0.135	***
LGDPPC5							-0.006	
Constant	-0.147	***	-0.150	***	-0.192	***	-0.196	***
Leads	4		1		1		2	
Lags	1		1		1		1	
AIC	-465		-460		-511		-567	
BIC	-412		-395		-433		-445	

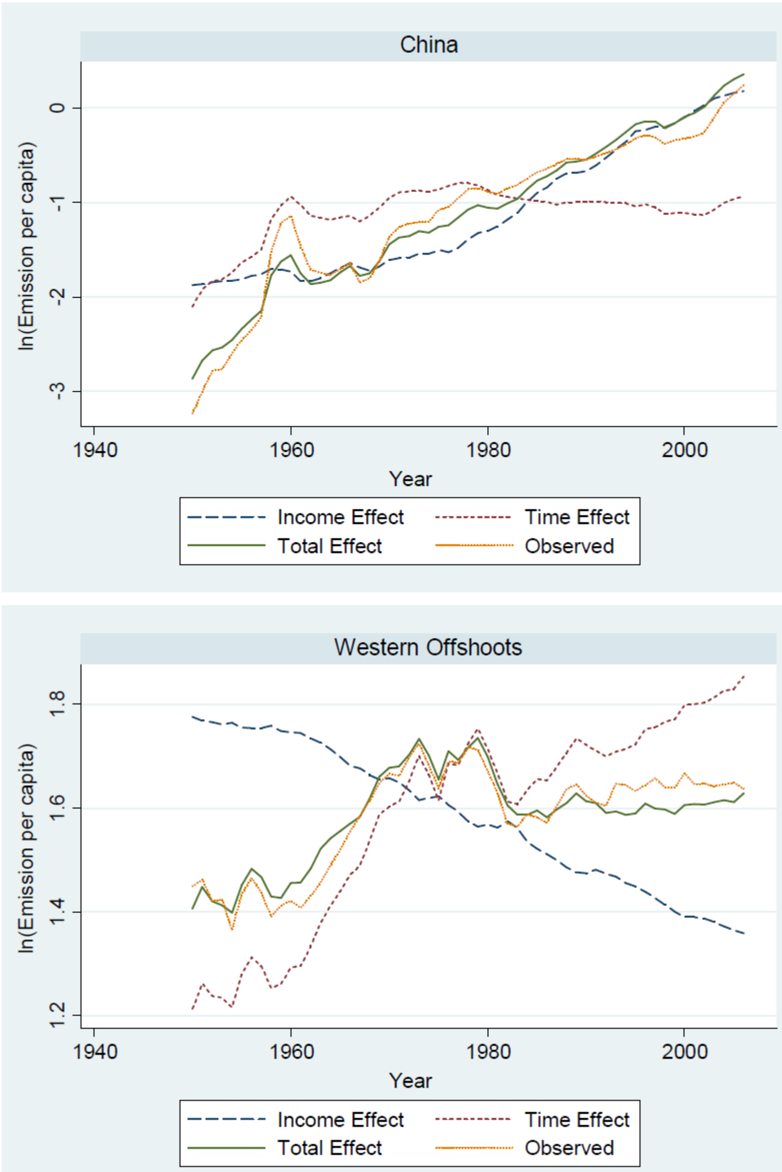
(*) null hypothesis can be rejected at 1% significance level.
(**) null hypothesis can be rejected at 5% significance level
(***) null hypothesis can be rejected at 10% significance level

the estimated change in emission per capita in a specific region for a given period, due to the given change in income or time (the income effect and time effect, respectively). The total effect is simply the sum of the income effect and the time effect. A first noteworthy point in Figure 2.18 is that, for China, the time effects, which represent the technological factors and the changes in sectoral composition, play a minor role in the total effect, and just the opposite for Western Offshoots. Secondly, the total effect is mostly increasing for both regions, and hence there is no evidence towards a decline in overall emissions. Thirdly, the income effect is positive for China, but negative for Western Offshoots. The U-shaped income-emission relation predicted by the EKC hypothesis in terms of income is observed in only the economically developed regions, namely Western Offshoots and Western Europe (not presented). Therefore, the income effects are responsible for the inverted U-shaped relation under homogeneity. However, this prediction is at odds to the interpretation of income as a scale effect. Firstly, time effects for these regions are estimated to be positive. This means that in these regions factors related to technological and sectoral composition changes contribute positively to the rising emissions. Secondly, for these two regions, the total effect is positive in most of the periods. That is, the total effect is driven by the time effect rather than the income effect. However, one would expect that in these regions the changes in emissions should be mainly dominated by the income related factors. This casts doubts on the supportive evidence for the EKC hypothesis from the estimations under homogeneity which might be a too strong assumption in our case.

2.B.1. Cointegration

The first and second generation panel unit root tests strongly suggest that our series are I(1) processes. Therefore, unless there is cointegration for the proposed relations, we might face a spurious regression problem when we regress emissions on income. Testing for cointegration in macro-panels is complicated by the presence of cross-sectional dependence and cross unit cointegration. For example, when modeling the cross-sectional dependence with a common factor structure, Gengenbach et al. (2006) suggest to test for cointegration among the idiosyncratic components of the variables, only if the common factors are stationary. Furthermore, if the common factors are also I(1) processes, then one needs to verify a cointegrating relationship among the factors. This procedure

Figure 2.18: Income and Time Effects of the Pairwise Differencing – DOLS Estimation with Fixed Effects under Full Homogeneity



requires a PANIC analysis which might be appropriate in our case due to the small time dimension. However, as discussed in Banerjee et al. (2004), these complications can be handled with a system-based cointegration test when the cross-sectional dimension is small relative to the time dimension, which is the case in our data set. A further complication in the EKC context is the presence of nonlinear transformations of the nonstationary variables. As noted by Wagner (2008), this requires a different asymptotic theory than the present panel cointegration tests.

In this section, we investigate the validity of the estimation strategies by conducting several cointegration tests. In a panel setting, cointegrating relationships can be present across cross-section units for the same variable or across variables measured on the same cross-section unit. More precisely, consider an observation z_{ijt} where $i = 1, \cdots, N$ is the cross-section unit, $j = 1, \cdots, K$ denotes K different variables, and finally $t = 1, \cdots, T$ is the time dimension. Assume that for any i and j, the series z_{ijt} is integrated of order one. Then a cointegrated relationship may exist for any combinations of $N \times K$ series. The first group of cointegration tests ignores the presence of cross-unit cointegration and tests for the presence of cointegrating relationships among the variables for each cross-sectional unit. These tests are residual-based tests and they can be considered as an extension of the first generation unit root tests, applied to the residuals of a regression among the panel variables. We will consider Kao (1999) and Pedroni (2004) as residual based tests. A second group of tests we perform are based on a system approach which allows for cross-unit cointegration. These are Maddala and Wu (1999) and Westerlund (2007). We apply these tests to the full panel, including all the regions (thus, also, India, Other Asia, Western Europe, Eastern Europe, Former USSR, Africa, and Latin America).

Residual based tests

The Kao (1999) test is residual based and ignores possible cross unit cointegration. The following model is considered:

$$
\begin{aligned}
y_{it} &= \delta_i' d_{it} + x_{it}' \beta + u_{it}, \\
\varepsilon_{it} &= \Delta x_{it} - E(\Delta x_{it}).
\end{aligned}
$$

The dependent variable y_{it} is the logarithm of emission per capita. The $K-1$ variable in x_{it}, which in our case are the logarithm of GDP per capita and its higher order terms. These are (assumed to be) $I(1)$ such that ε_{it} is a white noise. In addition, it is assumed that x_{it} is not cointegrated across i. However, we allow the error terms u_{it} and ε_{it} to be correlated. Finally, d_{it} denotes the deterministic component (whether there is a constant, trend, or none). The null of no cointegration amounts to testing the null of a unit root in u_{it} which is conducted in an ADF form regression like:

$$\Delta\hat{u}_{it} = \alpha + \sigma\hat{u}_{it-1} + \sum_{j=1}^{p} \theta_j \Delta\hat{u}_{i,t-1} + \upsilon_{it},$$

where $\hat{u}_{it}$ denotes the estimated u_{it}. The ADF coefficient is assumed to be homogeneous and therefore this test can be seen as an extension of the LLC panel unit root test. So, the null hypothesis is formalized as $H_0 : \sigma = 0$ against the alternative $H_0 : \sigma < 0$.

In Table 2.29 the results of the Kao test are presented for the emissions and the GDP series and for different polynomial specifications of the hypothesized relationship. For all specifications, the results indicate the presence of a cointegrating relationship.

While Kao's test is an extension of the LLC-test, the Pedroni (2004) tests are an extension of the IPS panel unit root test with a heterogeneous alternative. Therefore, the model allows for different cointegrating vectors across cross section units. That is, the parameter σ is allowed to be individual specific. In Table 2.30 results of the four tests proposed in Pedroni (2004) are presented. Each of them is applied to different specifications of the deterministic data generating component and different specifications of the alternative hypothesis. The rejection of the null of no cointegration is very rare. It seems that allowing for heterogeneous cointegrating vectors changes the results substantially. These tests do not account for cross-sectional dependence. Therefore, we apply the same tests for the demeaned variables by subtracting the cross-sectional averages. The findings are presented in Tables 2.31 and 2.32, which are similar to the previous results.

We also need to apply the same tests to the pairwise differences of the variables in order to test for the validity of the pairwise differencing estimation. For the case where we assumed heterogeneous income effects, the results are presented in Tables 2.33 and 2.34. The Kao test again indicates a cointegrating relationship for all polynomial specifications. Also, the Pedroni test gives results towards a cointegrating relationship.

Table 2.29: Kao (1999) Cointegration Test for logarithm of Emission and GDP series

Series	p-values
Linear eequation	0.0044
Quadratic equation	0.0007
Third order polynomial	0.0006
Fourth order polynomial	0.0011
Fifth order polynomial	0.0010

Table 2.30: Pedroni (2004) Cointegration Test for logarithm of Emission and GDP series

	None	constant	constant and trend
Linear equation			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.982	0.203	**0.019**
Panel rho-Statistic	0.744	0.398	0.112
Panel PP-Statistic	0.000	0.102	0.045
Panel ADF-Statistic	0.007	0.345	0.006
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.988	0.897	0.877
Group PP-Statistic	0.000	0.882	0.904
Group ADF-Statistic	0.049	0.984	0.964
Quadratic equation			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.984	0.043	**0.170**
Panel rho-Statistic	0.920	0.490	0.957
Panel PP-Statistic	0.809	0.425	0.844
Panel ADF-Statistic	0.942	0.123	0.174
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.985	0.430	0.892
Group PP-Statistic	0.263	0.430	0.878
Group ADF-Statistic	0.936	0.362	0.925
Third Order Polynomial			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.993	0.039	**0.063**
Panel rho-Statistic	0.980	0.589	0.934
Panel PP-Statistic	0.814	0.559	0.930
Panel ADF-Statistic	0.984	0.262	0.186
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.970	0.769	0.879
Group PP-Statistic	0.689	0.709	0.723
Group ADF-Statistic	0.965	0.213	0.633
Fourth order polynomial			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.999	0.150	0.149
Panel rho-Statistic	0.994	0.833	0.988
Panel PP-Statistic	0.994	0.820	0.994
Panel ADF-Statistic	1.000	0.285	0.620
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.994	0.688	0.945
Group PP-Statistic	0.938	0.647	0.935
Group ADF-Statistic	1.000	0.060	0.834
Fifth order polynomial			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	1.000	0.429	0.300
Panel rho-Statistic	0.997	0.966	0.997
Panel PP-Statistic	0.991	0.963	0.989
Panel ADF-Statistic	0.999	0.655	0.629
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.988	0.931	0.997
Group PP-Statistic	0.921	0.841	0.989
Group ADF-Statistic	0.984	0.183	0.952

Table 2.31: Kao (1999) Cointegration Test by Demeaning the logarithm of Emission and GDP series

Series	p-values
Linear equation	0.0026
Quadratic equation	0.0032
Third order polynomial	0.0031
Fourth order polynomial	0.0108
Fifth order polynomial	0.0141

Despite the presence of conflicting results, for the quadratic and cubic specifications most of the Pedroni tests reject the null of no cointegration as long as an intercept, or a deterministic trend, is included.

System Based Cointegration Tests

The framework suggested by Pedroni (2004) as well as Kao (1999) can be restrictive in our case, since for both the GDP and the emission series one may expect to find some long-term relationship across cross-sectional units. This section considers system based approaches which allow for cross-unit cointegration.

Maddala and Wu (1999) apply the Fisher et al. (1970) technique to combine individual unit root tests in order to create a combined Fisher-type panel unit root test. By applying the same idea, they suggest a cointegration test for panels. It is a multivariate likelihood-ratio analysis and it relies on the VAR representation of the variables. The null is "there are at most r cointegrating relationships" or "there are r cointegrating relationships" against the alternative that "there are more than r cointegrating relationships." The test is applied for an increasing number of r, until the null cannot be rejected. Results are presented in Table 2.35. For all series, both the trace-test and the max-eigen value test indicate the presence of a cointegrating relationship. Another finding is that in some cases the cointegration matrix is found to be of full rank which implies stationarity for the series subject to test. The results for the demeaned variables are presented in Table 2.36, which are similar.

For the pairwise differencing regression with heterogeneous income effects, there are similar findings presented in Table 2.37. Mostly, the test finds at least one cointegrating relationship, but sometimes indicates a full rank correlation matrix.

Westerlund (2007) provides several tests for the null of no cointegration. The distinghuishing feature of these tests is that they rely on an error-correction representation

Table 2.32: Pedroni (2004) Cointegration Test by Demeaning the logarithm of Emission and GDP series

	None	constant	constant and trend
Linear equation			
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$		
Panel v-Statistic	0.963	0.842	**0.392**
Panel rho-Statistic	0.777	0.934	0.900
Panel PP-Statistic	0.680	0.924	0.851
Panel ADF-Statistic	0.801	0.919	0.929
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$		
Group rho-Statistic	0.988	0.985	0.971
Group PP-Statistic	0.353	0.970	0.954
Group ADF-Statistic	0.563	0.966	0.976
Quadratic equation			
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$		
Panel v-Statistic	0.963	0.835	**0.560**
Panel rho-Statistic	0.935	0.942	0.972
Panel PP-Statistic	0.978	0.754	0.974
Panel ADF-Statistic	0.988	0.726	0.992
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$		
Group rho-Statistic	0.998	0.999	0.995
Group PP-Statistic	0.999	0.999	0.992
Group ADF-Statistic	0.995	0.999	0.996
Third Order Polynomial			
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$		
Panel v-Statistic	0.983	0.968	**0.562**
Panel rho-Statistic	0.880	0.993	0.912
Panel PP-Statistic	0.904	0.998	0.830
Panel ADF-Statistic	0.181	0.993	0.743
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$		
Group rho-Statistic	0.998	0.999	0.992
Group PP-Statistic	0.989	0.999	0.951
Group ADF-Statistic	0.791	0.999	0.829
Fourth order polynomial			
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$		
Panel v-Statistic	0.996	0.965	0.751
Panel rho-Statistic	0.989	0.992	0.896
Panel PP-Statistic	0.997	0.946	0.757
Panel ADF-Statistic	0.328	0.945	0.707
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$		
Group rho-Statistic	0.999	0.999	0.984
Group PP-Statistic	0.999	0.999	0.895
Group ADF-Statistic	0.925	0.997	0.796
Fifth order polynomial			
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$		
Panel v-Statistic	0.998	0.967	0.896
Panel rho-Statistic	0.975	0.997	0.978
Panel PP-Statistic	0.992	0.940	0.943
Panel ADF-Statistic	0.332	0.946	0.879
	H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$		
Group rho-Statistic	0.996	0.999	0.998
Group PP-Statistic	0.999	0.999	0.987
Group ADF-Statistic	0.959	0.997	0.925

Table 2.33: Kao (1999) Cointegration Test for the Pairwise Differencing Estimations with Heterogenous Income Effects

Series	p-values
Quadratic equation	0.0034
Third order polynomial	0.0019
Fourth order polynomial	0.0002
Fifth order polynomial	0.0005

Table 2.34: Pedroni (2004) Cointegration Test for the Pairwise Differencing Estimations with Heterogenous Income Effects

	None	constant	constant and trend
Linear equation			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.243	0.011	**0.000**
Panel rho-Statistic	0.215	0.376	0.429
Panel PP-Statistic	0.073	0.244	0.119
Panel ADF-Statistic	0.090	0.015	0.000
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.959	0.590	0.020
Group PP-Statistic	0.512	0.329	0.000
Group ADF-Statistic	0.559	0.114	0.000
Quadratic equation			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.223	0.000	**0.019**
Panel rho-Statistic	0.793	0.075	0.330
Panel PP-Statistic	0.752	0.017	0.106
Panel ADF-Statistic	0.428	0.000	0.000
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.393	0.000	0.011
Group PP-Statistic	0.521	0.000	0.000
Group ADF-Statistic	0.418	0.000	0.000
Third Order Polynomial			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.528	0.000	**0.037**
Panel rho-Statistic	0.981	0.141	0.607
Panel PP-Statistic	0.995	0.051	0.344
Panel ADF-Statistic	0.970	0.000	0.013
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.443	0.001	0.088
Group PP-Statistic	0.600	0.000	0.000
Group ADF-Statistic	0.989	0.022	0.471
Fourth order polynomial			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.672	0.035	0.283
Panel rho-Statistic	0.956	0.644	0.893
Panel PP-Statistic	0.958	0.404	0.736
Panel ADF-Statistic	0.741	0.031	0.020
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.328	0.036	0.174
Group PP-Statistic	0.367	0.000	0.000
Group ADF-Statistic	0.412	0.095	0.065
Fifth order polynomial			
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i = \sigma < 0$ *and* $\delta_{i1} \in R$			
Panel v-Statistic	0.846	0.089	0.491
Panel rho-Statistic	0.987	0.746	0.961
Panel PP-Statistic	0.984	0.518	0.866
Panel ADF-Statistic	0.866	0.034	0.084
H0: $\sigma_i = \sigma = 0$ *and* $\delta_{i1} = 0$; **H1:** $\sigma_i < 0$ *and* $\delta_{i1} \in R$			
Group rho-Statistic	0.586	0.258	0.687
Group PP-Statistic	0.206	0.008	0.028
Group ADF-Statistic	0.391	0.248	0.328

Table 2.35: Johansen Fisher Panel Cointegration Test for the logarithm of Emission and GDP series

Degree of Polynomial Specifications	one	two	three	four	five
A					
Trace Test	Full Rank	2	Full Rank	Full Rank	Full Rank
Max Eigenvalue Test	Full Rank	2	Full Rank	Full Rank	Full Rank
B					
Trace Test	Full Rank	Full Rank	Full Rank	Full Rank	Full Rank
Max Eigenvalue Test	Full Rank	Full Rank	Full Rank	Full Rank	Full Rank
C					
Trace Test	Full Rank	2	3	Full Rank	5
Max Eigenvalue Test	Full Rank	2	3	Full Rank	5
D					
Trace Test	1	2	Full Rank	Full Rank	Full Rank
Max Eigenvalue Test	1	2	Full Rank	Full Rank	Full Rank
E					
Trace Test	0	Full Rank	3	4	5
Max Eigenvalue Test	0	1	3	4	5

5% siginificance level
A: no trend or intercept in CE or VAR
B: intercept in CE - no intercept in VAR
C: intercept in CE and VAR
D: intercept and trend in CE - no trend in VAR
E: intercept and trend in CE - linear trend in VAR

Table 2.36: Johansen Fisher Panel Cointegration Test by Demeaning Data the logarithm of Emission and GDP series

Degree of Polynomial Specifications	one	two	three	four	five
A					
Trace Test	0	2	3	4	5
Max Eigenvalue Test	0	2	3	2	3
B					
Trace Test	0	2	2	3	4
Max Eigenvalue Test	0	2	3	2	4
C					
Trace Test	0	1	2	3	4
Max Eigenvalue Test	0	1	2	2	4
D					
Trace Test	0	2	3	3	4
Max Eigenvalue Test	0	1	2	2	4
E					
Trace Test	Full Rank	1	2	2	4
Max Eigenvalue Test	Full Rank	1	2	2	3

5% siginificance level
A: no trend or intercept in CE or VAR
B: intercept in CE - no intercept in VAR
C: intercept in CE and VAR
D: intercept and trend in CE - no trend in VAR
E: intercept and trend in CE - linear trend in VAR

Table 2.37: Johansen Fisher Panel Cointegration Test for the Pairwise Differencing Estimations with Heterogenous Income Effects

Degree of Polynomial Specifications	one	two	three	four	five
A					
Trace Test	1	1	3	4	5
Max Eigenvalue Test	1	1	3	4	5
B					
Trace Test	1	2	Full Rank	Full Rank	Full Rank
Max Eigenvalue Test	1	2	Full Rank	Full Rank	Full Rank
C					
Trace Test	Full Rank	Full Rank	2	3	5
Max Eigenvalue Test	Full Rank	1	2	3	5
D					
Trace Test	1	2	3	4	Full Rank
Max Eigenvalue Test	1	2	3	4	4
E					
Trace Test	Full Rank	Full Rank	Full Rank	3	4
Max Eigenvalue Test	Full Rank	1	1	3	4

5% siginificance level
A: no trend or intercept in CE or VAR
B: intercept in CE - no intercept in VAR
C: intercept in CE and VAR
D: intercept and trend in CE - no trend in VAR
E: intercept and trend in CE - linear trend in VAR

of the data. In an error correction model, an insignificant error-correction coefficient implies the null of no cointegration. Westerlund (2007) highlights the important difference between the residual based tests and the proposed error-correction based tests. The former one relies on a restriction which the author refers to as the common factor restriction which is possibly invalid. On the other hand, error correction based tests assume weak exogeneity. It is mentioned that the test choice depends on a trade-off between these restrictions.

The data generating process is assumed to be as follows,

$$\begin{aligned} y_{it} &= \phi_{1i} + \phi_{2i}t + z_{it}, \\ x_{it} &= x_{i,t-1} + v_{it}. \end{aligned}$$

Here, x_{it} is a vector of variables which are modeled as random walk processes. On the other hand, the dependent variable y_{it} is composed of a deterministic part and a stochastic part, denoted by z_{it}. The stochastic term is modeled as a conditional error correction model as follows:

$$\alpha_i(L)\Delta z_{it} = \alpha_i(z_{i,t-1} - \beta_i' x_{i,t-1}) + \gamma_i(L)' v_{it} + e_{it},$$

where $\alpha_i(L) = 1 - \sum_j \alpha_{ij} L^j$ and $\gamma_i(L) = \sum_j \gamma_{ij} L^j$ are polynomials in terms of the lag operators. The resulting conditional error-correction model is,

$$\alpha_i(L)\Delta y_{it} = \delta_{1i} + \delta_{2i}t + \alpha_i(y_{i,t-1} - \beta_i^{'}x_{i,t-1}) + \gamma_i(L)'\upsilon_{it} + e_{it}. \quad (2.12)$$

In order to see the common factor restriction, one subtracts $\alpha_i(L)\beta_i^{'}\upsilon_{it}$ from both sides of equation (2.12),

$$\Delta(y_{it} - \beta_i^{'}x_{i,t}) = \delta_{1i} + \delta_{2i}t + \alpha_i(y_{i,t-1} - \beta_i^{'}x_{i,t-1}) + (\gamma_i(L) - \alpha_i(L)\beta_i^{'}\upsilon_{it}) + e_{it}. \quad (2.13)$$

Westerlund's residual based approach tests the null of no cointegration by testing $\alpha_i = 0$. The regression on which the tests are based can be stated by rewriting equation (2.13),

$$\Delta y_{i,t} = \delta_{1i} + \delta_{2i}t + \alpha_i y_{i,t-1} + \lambda_i^{'}x_{i,t-1} + \sum_{j=1}^{p}\alpha_{ij}\Delta y_{i,t-j} + \sum_{j=1}^{p}\gamma_{ij}\Delta x_{i,t-j} + e_{it}.$$

There are four tests proposed by Westerlund (2007): The first two test statistics (Gt and Ga) are group-mean statistics for which the alternative hypothesis is presence of at least one cointegrating relationship in the panel. The second group of tests provides panel based statistics (Pt and Pa) for which the alternative hypothesis is the presence of cointegration for all cross-section units. The tests indicated with a t subscript are simple t-tests and the ones with a p subscript are averaging $T\hat{\alpha}_i/\alpha(L)$ over cross-sectional units.

The proposed tests assume cross-sectional independence. In order to account for the cross-sectional dependence, we performed the tests with bootstrapping, as suggested by Westerlund (2007).

The proposed tests also allow for testing for the presence of individual intercepts and individual trends. The results are presented in Table 2.38 for the baseline estimations. The null of no cointegration is rejected in only one case for which the null is that there is at least one cointegrating relationship. The results for the demeaned data without bootstrapping is presented in Table 2.39. Again, the null of no cointegration is not rejected by almost all the tests.

For the pairwise differencing regression with heterogeneous income effects, the results presented in Table 2.40 are slightly different. As long as a constant and/or trend is included, the likeliness of a cointegrating relation increases. Besides, for the higher order polynomials likeliness of a cointegrating relation decreases. For the quadratic equation, most of the tests by Westerlund indicates a cointegrating relationship at the 10% significance level.

Table 2.38: Westerlund Cointegration Test p-values with Bootstrapping (100) for logarithm of Emission and GDP series

Order of polynomial	one	two	three	four	five
Constant & Trend					
Gt	0.740	0.780	0.740	0.830	0.960
Ga	0.820	1.000	1.000	1.000	0.990
Pt	0.770	0.740	0.760	1.000	0.920
Pa	0.480	0.640	0.770	0.890	0.930
Constant					
Gt	0.730	0.550	0.400	0.710	0.870
Ga	0.190	0.460	0.520	0.700	0.910
Pt	0.320	0.170	0.050	0.080	0.380
Pa	0.280	0.140	0.000	0.040	0.690
None					
Gt	0.090	0.980	0.940	0.920	0.880
Ga	0.880	1.000	0.980	1.000	0.990
Pt	0.410	0.720	0.500	0.690	0.850
Pa	0.620	0.770	0.850	0.890	0.990

Table 2.39: Westerlund Cointegration Test p-values without / with Bootstrapping (100) for Demeaned Data

Order of polynomial	one	two	three	four	five
Constant & Trend					
Gt	0.970/0.960	0.969/0.870	0.992/0.000	0.987/0.000	0.996/0.000
Ga	0.963/0.890	0.989/0.890	0.999/0.000	1.000/0.000	1.000/0.000
Pt	0.347/0.490	0.428/0.420	0.773/0.000	0.744/0.000	0.969/0.000
Pa	0.139/0.170	0.277/0.180	0.719/0.000	0.925/0.000	0.989/0.000
Constant					
Gt	0.793/0.710	0.925/0.850	0.979/0.000	1.000/0.000	0.999/0.000
Ga	0.938/0.870	0.979/0.920	0.997/0.000	1.000/0.000	1.000/0.000
Pt	0.320/0.190	0.046/0.200	0.226/0.000	0.885/0.000	0.957/0.000
Pa	0.280/0.250	0.448/0.320	0.810/0.000	0.997/0.000	0.995/0.000
None					
Gt	0.006/0.000	0.349/0.340	0.714/0.000	0.938/0.000	0.710/0.000
Ga	0.779/0.550	0.964/0.920	0.996/0.000	1.000/0.000	1.000/0.000
Pt	0.001/0.040	0.037/0.230	0.223/0.000	0.690/0.000	0.549/0.000
Pa	0.050/0.190	0.404/0.404	0.782/0.000	0.944/0.000	0.958/0.000

Table 2.40: Westerlund Cointegration Test for Pairwise Differencing (Heterogeneity) (p-values with Bootstrapping (100))

Order of polynomial	one	two	three	four	five
Constant & Trend					
Gt	0.170	0.070	0.190	0.130	0.660
Ga	0.100	0.020	0.260	0.710	0.960
Pt	0.170	0.240	0.470	0.800	0.910
Pa	0.190	0.080	0.440	0.800	0.910
Constant					
Gt	0.940	0.030	0.160	0.520	0.600
Ga	0.710	0.000	0.070	0.530	0.930
Pt	0.270	0.110	0.180	0.610	0.530
Pa	0.160	0.010	0.140	0.470	0.500
None					
Gt	0.380	0.350	0.310	0.250	0.640
Ga	0.470	0.460	0.530	0.750	0.950
Pt	0.050	0.360	0.500	0.480	0.610
Pa	0.100	0.320	0.510	0.430	0.770

2.B.2. Nonparametric Confidence Intervals

We use the smooth-backfitting Nadaraya-Watson estimator following Mammen et al. (1999), and as explained in Nielsen and Sperlich (2005).

We have a regression model $E(Y|X = x) = m(x)$, $Y \in \mathbb{R}$, $X \in \mathbb{R}^d$, where x stands for a fixed point and where we can model this regression additively as :

$$m(x) = m_0 + \sum_{j=1}^{d} m_j(x_j)\,,$$

where $m_0 = E(Y)$. Identification requires $E\{(m_j(X_j)\} = 0$ for all $j > 0$. Denoting the sample size with n, and the the bandwidth with h, such that $n^{1/5}h \to c_h$ for a constant c_h. Mammen et al. (1999) show that following convergence holds:

$$n^{2/5}\begin{pmatrix} \hat{m}_1(x_1) - m_1(x_1) \\ \vdots \\ \hat{m}_d(x_d) - m_d(x_d) \end{pmatrix} \to N\left\{\begin{pmatrix} c_h^2\beta_1(x_1) \\ \vdots \\ c_h^2\beta_d(x_d) \end{pmatrix}, \begin{pmatrix} v_1(x_1) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & v_d(x_d) \end{pmatrix}\right\},$$

where $\hat{m}_j$ is the smooth backfitting estimate of the j'th additive term, and β_j is a function given by

$$(\beta_0, ..., \beta_d) = \arg\min_{\beta_0,..,\beta_d}\left[\int \{\beta(x) - \beta_0 - \beta_1(x_1) - ... - \beta_d(x_d)\}^2 p(x)dx\right],$$

where $p(x)$ is the multidimensional kernel density function of X. Our estimated income effects corresponds to individual $m_j(x_j)$'s. Therefore, the confidence intervals are calculated as:

$$CI_j(x_j) = \hat{m}_j(x_j) \pm 1.96\sqrt{v_j(x_j)},$$

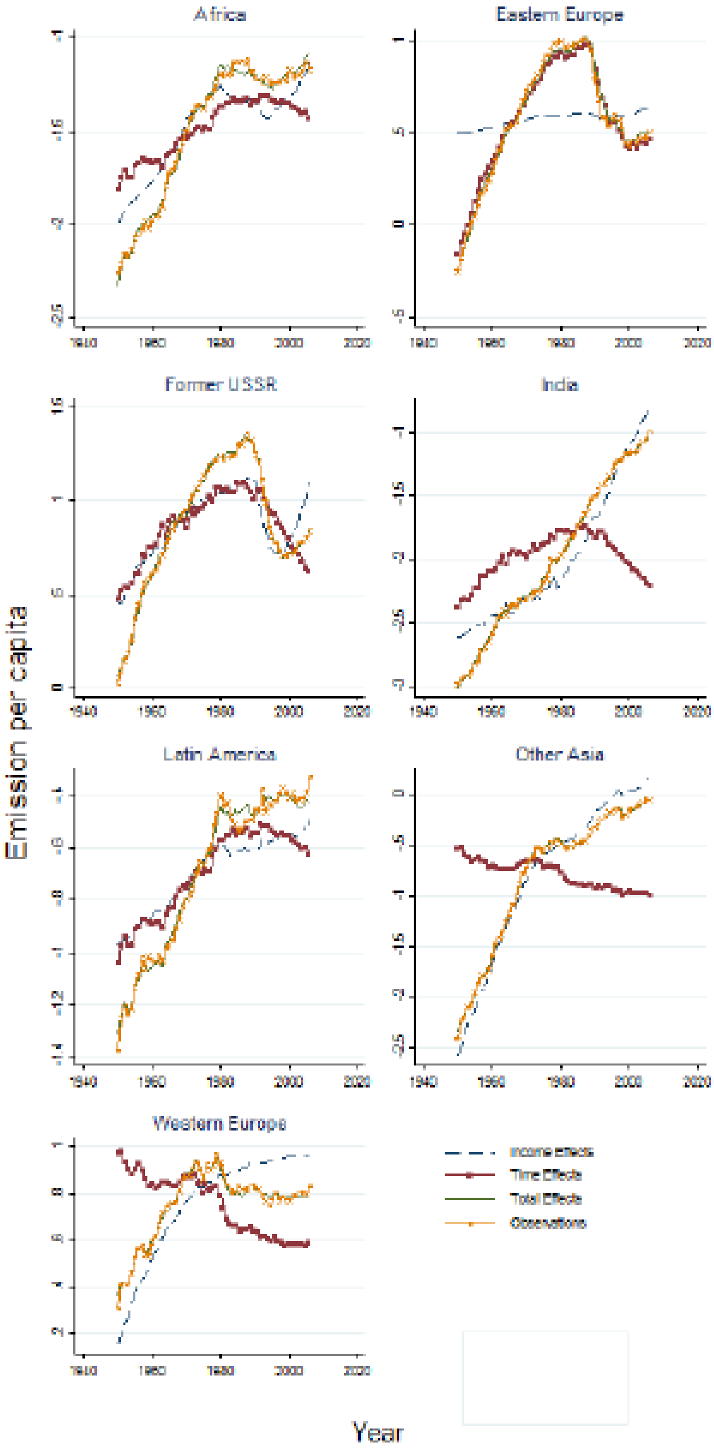
where $v_j(x_j) = (c_K/c_h)\sigma_j^2(x_j)/p_j(x_j)$. The term c_K is equal to $1/(2\sqrt{\pi})$, we take c_h equal to one, $\sigma_j^2(x_j)$ is the variance of residuals conditional upon $X_j = x_j$, and $p_j(x_j) = \int p(x)dx_{-j}$ is the marginal density function.

2.B.3. Some estimation results for the other regions

Figure 2.19 shows the estimations of the non-parametric non-stationary approach with pairwise differencing of the other regions. We use the pairs according to Melenberg et al.

(2011): Africa is coupled to Latin America, Eastern Europe is coupled to Former USSR, Former USSR is coupled to India, India is coupled to Former USSR, Latin America is coupled to Africa, Other Asia is coupled to Western Offshoots, and Western Europe is coupled to Western Offshoots.

Figure 2.19: Estimation Results Other Regions



Chapter 3

Pairwise Differencing Forecast of Global Carbon Dioxide Emissions: China vs. Time Effects

3.1. Introduction

Considered as one of the most important factors leading to global warming, carbon dioxide (CO_2) emissions as a result of economic activity lies at the core of the debates on climate change. How much CO_2 will be emitted in the future is important to the international community to understand the urgency and stringency of the measures that should be taken, and central to these discussions is the uncertainty of future emissions. The literature on the modeling of CO_2 emissions is predominated by structural models, in which some structural parameters, (such as population growth, income growth, and technological change) are chosen by expert judgements. This subjective uncertainty is the main drawback of these models in forecasting future emissions. For example, in the Special Report on Emission Scenarios (SRES) by the Intergovernmental Panel on Climate Change (IPCC), despite the underlying "no change in policy" assumption (see IPCC (2000) and IPCC (2008)), future forecasts of greenhouse gas emissions range from a level that is over five times larger than the current level to a reduction by 2100, depending on the subjective uncertainty. Therefore, probabilistic business-as-usual forecasts as a benchmark to the structural models are crucially important (Schmalensee et al., 1998).

Although reduced form econometric modeling is a convenient tool for the purpose of business-as-usual forecasts, and widely used in order to investigate the in-sample functional relation between CO_2 emissions and Gross Domestic Production (GDP) per capita in the Environmental Kuznets Curve (EKC) literature, attempts to make future forecasts depending on this relationship is scarce. The main difficulty is that the estimated in-sample functional relations are generally non-linear (such as second or third order polynomial functions) which potentially lead to implausible future forecasts such as explosive or zero emissions even in the near future. In this paper we argue that there is a fundamental endogeneity problem in the reduced form EKC models, stemming from the insufficiency of time trends to proxy potentially mitigating effects (time effects) such as technology, industrial composition, preferences, policy, and so on. We deal with this problem by incorporating the pairwise differencing approach proposed by Vollebergh et al. (2009) for a stationary context, and further developed by Sen et al. (2014b) to a non-stationary context. We extend this approach by modeling the time effects in order to be able to forecast regional and global level CO_2 emissions up to 2050 by extrapolating the income and the modeled time effects. In Vollebergh et al. (2009) and Sen et al. (2014b) income effects are identified independent of the identification problem of the time effects, under a common time trend assumption between pair regions, but the time effects are not really modeled. In this paper, we further model the time effects by treating them as a residual data. That is, our estimation strategy identifies both pure income related emission pathways (scale effects) and fully flexible time effects which are expected to follow monotonic trends. While the income (scale) effects should always be increasing in income per capita, time effects are expected to be decreasing for developed economies and increasing for developing economies (unless there is a trend break). Extrapolating these monotonic trends is easier compared to extrapolating the potentially nonlinear total effects.

By disaggregating the total effects into its components, income and time effects, and extrapolating these separate effects under the assumption of no trend breaks, we are able to investigate some important questions regarding the international negotiations in an effort to reduce CO_2 emissions: Firstly, does the rapid carbon intensive industrialization, experienced in the developing regions, like China, constitute the major future threat? Secondly, are the pollution compensating factors, like the advancement in green technologies in the developed world, sufficient to reduce future emissions at both the regional and the global level?

Our main findings are as follows: Firstly, according to our extrapolations, without a trend break, global CO_2 emissions will rise steadily up to 2050, where some optimistic non-intervention IPCC scenarios indicating a U-turn in our forecast period stay outside our 95% confidence interval. Secondly, our extrapolations imply that the income effects are rising for all regions, and that the time effects of the developed regions partially offset the rising income effects. This leads to a stablilization of the total effects in the developed regions. However, the global level extrapolations show that the negative time effects of the developed regions are far from creating a slow-down in the global total effects, so an inverted U-shaped relation as suggested by the EKC hypothesis is not likely to be observed at the global level future emissions, given our business-as-usual scenario. These results are fully in line with the theoretical arguments in the EKC literature that environmental degradation, due to a growing scale of the economy, is mitigated as the economy grows above a threshold level which induces technological change and sectoral composition towards a more environmental friendly point. Thirdly, the income effect of China is a strong contributor to the global emissions, reflecting their recent high growth rates. Moreover, the estimated time effects of China are positive, potentially reflecting their switch to a coal-based energy input mix, in combination with a shift to industrial production. However, any scenarios excluding the "China effect" does not change the global picture. That is, our results indicate that the source of both current and future growth in global emissions is not mainly China, but the insufficient progress in the green technologies of developed regions.

This paper is closely related to the EKC literature, initiated by Grossman and Krueger (1991). The early literature focuses on various indicators for environmental degradation, and analyzes many sub-samples of countries or regions, by employing parametric reduced form models (Shafik and Bandyopadhyay, 1992; Selden and Song, 1994; Panayotou, 1993; Horvath, 1997; Komen et al., 1997; De Bruyn et al., 1998; Stern, 1998). As follow up of this early literature, the attention turned towards the econometric techniques employed in the early literature. First, the parametric estimation of the EKC relation is criticized for being simply a trial and error approach by testing a pre-specified functional form, and as a solution, non-parametric or semi-parametric econometric techniques were employed (Taskin and Zaim, 2000; Millimet et al., 2003; Azomahou et al., 2006). Second, the developments in non-stationary panel data estimation techniques necessitated to revise the earlier findings, by also taking into account the non-linear specification of a potentially non-stationary variable (Stern, 2004; Muller-Furstenberger and Wagner, 2007; Wagner, 2008; Galeotti et al., 2009). Third, Dijkgraaf and Vollebergh

(2005) showed that the homogeneous parameter assumption in panel estimations is too strong. Allowing heterogeneity, Martinez-Zarzoso and Bengochea-Morancho (2004) used a pooled mean estimator. Fourth, Vollebergh et al. (2009) argued that the empirical EKC literature suffers from a fundamental problem due to the identification of the time effects. They proposed pairwise differencing as a remedy but not allowing for non-stationarity. In this paper, our in-sample estimations are based on the extension of the method proposed in Sen et al. (2014b), controlling for all these criticisms. Moreover, we extend Sen et al. (2014b) by updating the dataset with more recent data.

Our paper also directly contributes to the literature forecasting global emissions with reduced form models. While Holtz-Eakin and Selden (1995) use a quadratic specification, Schmalensee et al. (1998) use a flexible estimator. Auffhammer and Carson (2008) forecast the CO_2 emission pathways of China, by aggregating the provincial level forecasts. In a similar manner, Auffhammer and Steinhauser (2012) forecast US emissions, by focusing on model selection. Despite the difference that we present global level forecast, our forecasting strategy is similar in the sense that we also use regional forecasts to construct our global level forecasts. Similar to our approach of dismantling the total effects, disaggregation across regions can also improve the forecasting performance (Giacomini and Granger, 2004; Marcellino et al., 2003). None of the mentioned papers account for potential non-stationarity, parameter heterogeneity, and the endogeneity problem. Our approach allows us to figure out the driving force of the change in forecasted emissions, whether it is time effects or income effects of specific regions.

The remainder of this chapter is organized as follows. Section 2 describes the empirical strategies. Our dataset is described in section 3. In section 4, estimation and extrapolation results are analyzed. Finally, section 5 contains the conclusion.

3.2. Endogeneity Problem

Structural models in the natural science and engineering literature, such as the so-called integrated assessment models (IAMs) employed by IPCC, are based on the IPAT identity (Ehrlich et al., 1971; Commoner, 1972), where the impact (I) is decomposed into three multiplicative determinants: Population (P), affluence (A), and technology (T). This identity is formalized as follows:

$$I = P \times A \times T$$

Here affluence is generally proxied by a measure of economic activity scaled by population (GDP per capita), and $A \times T$ constitutes the total per capita impact. The IPAT equation may support many different points of view (see the survey by **?**), and although it is not explicitly stated in the literature, the reduced form EKC model is also a very flexible variant of the IPAT identity:

$$\log(I/P) = \log A + \log T + \varepsilon,$$

where ε is a stochastic error term. The common practice is to use CO_2 emission levels as a measure of impact (I), and modeling the impact of affluence (A) as a function of GDP per capita, rather than directly substituting GDP per capita for A.

Here, a series of problems arise due to the residual term T, which reflects not only the effect of technology, but possibly also other factors, such as changes in industrial composition or environmental policy (**?**). This term is difficult to measure at a macro level study, and generally interpreted as a residual term (Dietz and Rosa, 1994). In order to circumvent this problem, the EKC literature generally uses time dummies or a linear or quadratic time trend as a proxy for this composite effect. However, such an approach might not properly disentangle the scale effects from the effect of the composite term T. The reason is as follows: If the time effects, constituting the effect of factors other than income, are poorly proxied by the postulated time trends, then (If A is not tobe changed) part of their effect will be treated as a part of the error term. Since income and the other potentially mitigating factors are likely to be correlated, an endogeneity problem arises. Therefore, the estimated functional relation between emissions and GDP per-capita, not taking this endogeneity into account, might capture not only the scale effects, but possibly also the part of other factors including technology and industrial composition (see Auffhammer and Carson 2008; Vollebergh et al. 2009, and Sen et al., 2013). This leads to an omitted variable bias in the model parameters.[21] In line with this argument, the estimated functions might be non-linear, like inverted U-shaped or N-

[21]See Wooldridge (2010) for an extensive discussion of endogeneity problem in case of imperfectly proxied omitted variables.

shaped, while properly identified scale effects will not be decreasing in GDP per capita. As a result, forecasts depending on these functional forms might lead to counter-intuitive results, like explosive growth or zero emissions even in the near future. This is elaborated as the identification problem of the time effects by Vollebergh et al. (2009). It is argued that the imposed structure on the time effects (like linear or quadratic time trends) is consequential for the estimated functional form of the relation between emissions and income. Once the income effects are properly identified, they constitute the pure scale effects.

Such a disaggregation can also be very useful for extrapolation, since the potentially nonlinear total effect is decomposed into its determinants, which are more likely to have monotonic trends unless there is a structural break, and can be extrapolated easily. Forecasting the individual income and time effects, instead of the total effect can improve the forecasting performance (Lütkepohl, 1984; Lükepohl, 1987; Lütkepohl, 2006). In order to achieve this decomposition, we extend the pairwise differencing approach, first proposed by Vollebergh et al. (2009), and modified to account for the "non-linear specification of non-stationary covariates" in Sen et al. (2014b).

3.3. Empirical Strategy

In this section, we explain our empirical strategy. We start with a non-technical description of the in-sample estimations (see Sen et al. (2014b) for further details). Next, we describe the extrapolation procedure.

3.3.1. In-sample Estimation Strategy

The correlation between per-capita emissions and GDP, documented in the EKC literature, is basically a combination of the growing scale of the economy, structural changes in the composition of industries, and the extent to which such developments are affected by technological change or differences in resource availability. In order to gain insight into these structural trends, following the IPAT identity, the reduced form panel estimation technique postulates the general decomposition:

$$y_{it} = f(x_{it}, i) + \lambda(i, t) + \varepsilon_{it}, \tag{3.1}$$

where the logarithm of emissions per-capita, $y_{it} = \log(I_{it}/P_{it})$, of region (or country) i in year t is a combination of income effects, $f(x_{it}, i) = \log A_{it}$, and time effects, $\lambda(i,t) = \log(T_{it})$. Here, the income effect is modeled as a fully flexible function f of GDP per-capita, x_{it}, the time effects as a fully flexible function λ of time, and both effects are fully heterogeneous (i.e.., region specific). Proper identification of the scale effects requires to disentangle the effect of x_{it} from the (unobserved) time effects. The common approach is to assume that the heterogeneity can be fully captured by the fixed effects that are additively separable from f and λ, which leads to $y_{it} = \alpha_i + f(x_{it}) + \lambda(t) + \varepsilon_{it}$, where the income and time effects are assumed to be homogeneous across regions. Furthermore, the commonly applied parametric estimation methods postulate functional forms for f and λ, such as using some degree of polynomials or time dummies. However, as argued by Vollebergh et al. (2009), the choice of such functional forms are arbitrary to some degree, and raises a fundamental problem in the identification of the income effects. More specifically, the choice for λ is consequential for the estimated shape of f. This is a more crucial problem, if the goal is not just to test a postulated functional relation, but to identify the functional relation to make future forecasts. Indeed, the results in the EKC literature illustrate how problematic it is to obtain robust estimations of the long-term relationship between income and the environmental quality, even with comparable data sets. Therefore, it is important to be as flexible as possible, when specifying the time effects.

We start with the general form in equation (3.1), and do not impose any functional form for the time effects. Instead, by applying pairwise regional time differencing such that the time series data of the paired regions are substracted from eachother, we eliminate the common time effects of paired regions. Formally, consider the following reduced form relations for a pair of regions i and k:

$$\begin{aligned} y_{it} &= f_c(x_{i,t}, i) + \lambda_c(t) + \varepsilon_{c,it} \\ y_{kt} &= f_c(x_{k,t}, k) + \lambda_c(t) + \varepsilon_{c,kt} \end{aligned}$$

where $c = \{i, k\}$ indicates the pair. Here $\lambda_c(t)$ represents the common time effect, and $f_c(x_{i,t}, i)$ and $f_c(x_{k,t}, k)$ are the resulting region-specific income effects. Applying a pairwise differencing eliminates the common time effects leading to the following equation:

$$(y_{it} - y_{kt}) = f_c(x_{i,t}, i) - f_c(x_{k,t}, k) + (\varepsilon_{c,it} - \varepsilon_{c,kt}) \tag{3.2}$$

where we assume $E(\varepsilon_{c,it} - \varepsilon_{c,kt}|x_{it}, x_{kt}) = 0$. Therefore, this equation allows estimation of both $f_c(x_{i,t}, i)$ and $f_c(x_{k,t}, i)$ without imposing any functional restrictions on the time effects. Independent of the degree of the similarity of the time effects of the pair regions, the pairwise differencing is a more flexible approach in controlling for the common time effects compared to standard panel estimations which impose the assumption of equal time effects across all cross-sectional units. Moreover, its advantage in eliminating time effects increases, when the time effects of pair regions are more similar.

Different sets of pairs c correspond to different time effects and generate different income effects. The income effect in this set-up captures the change in the emissions not yet captured by the common time effect. In the sequel, we proceed under the assumption that for each region i there exists a set of pairs $c_i = (i, k_i)$ such that $\lambda_{c_i}(t)$ is the time effect of region i. Choosing the pair regions is key to the identification of the income effect in our approach. Our prior is that, combining two regions with similar time trends will result in a good fit, while combining two regions with different time trends will result in a bad fit. Based on this prior and on the basis of the in-sample fit of equation (3.2), for each region, we select a corresponding region with a similar time trend. This selection procedure is referred to as the "Goodness-of-Fit (GoF) prior". Any specification of the time effect, such as one being fixed and homogeneous across cross-sections, is also based on some prior. Our approach simply makes explicit, from the very beginning, that the empirical evidence on the the EKC relationship cannot be inferred automatically, but always depends upon one's prior (Heckman, 2000).

One can impose further restrictions on equation (3.2), depending on the employed estimation strategy. In Sen et al. (2014b), a wide range of estimations accounting for several problems are conducted. Firstly, it is shown that homogeneity of income effects across regions is a very strong assumption, driving the inverted U-shaped results in the literature. Dijkgraaf and Vollebergh (2005), by employing a formal test, reach the same conclusion. So, here, we assume full heterogeneity across regions, as well as pairs. Secondly, equation (3.2) can be estimated both parametrically and non-parametrically. Non-parametric techniques are superior by allowing a fully flexible estimation of functional forms. However, they might suffer from over-fitting and end-of sample biases, which deteriorates their usefulness in forecasting. Therefore, in this paper, we perform semi-parametric estimations of equation (3.2), by using polynomials up to the fifth order. In contrast to Vollebergh et al. (2009), we control for nonlinearity of non-stationary variables, by adopting the estimation strategy "efficient nonstationary nonlinear least squares" (ENNLS) suggested by Chang et al. (2001). Thus, we also take into account both the non-linearity and non-stationary properties of the variables. Indeed, Sen et al. (2014b) show that per capita emission and GDP series are potentially integrated of order one. [22]

Following the estimation of the income effects, the residual time effects can be obtained from:

$$\begin{aligned}
\lambda(t, i) + \varepsilon_{it} &= y_{it} - \hat{f}(x_{it}, i) \\
\lambda(t, k) + \varepsilon_{kt} &= y_{kt} - \hat{f}(x_{kt}, k),
\end{aligned}$$

where $\hat{f}(.)$ is the estimated income effects. Under the assumption that $\lambda(t) = \lambda(t, i) = \lambda(t, k)$, the difference between the two expressions is idiosyncratic. Vollebergh et al. (2009) suggest to callibrate $\lambda(t)$ as average of $y_{it} - \hat{f}(x_{it}, i)$ and $y_{kt} - \hat{f}(x_{kt}, k)$. Finally, the total effect is estimated as the sum of the estimated income and time effects.

This procedure reveals the time effects as residuals (i.e., observed minus estimated income effects). Therefore, we consider these revealed time effects as a series to be modeled econometrically. In this paper, we extend Vollebergh et al. (2009) and Sen et al. (2014b) by modeling the revealed time effects as univariate autoregressive integrated moving average (ARIMA) processes, possibly with deterministic trends.

In the EKC literature, an estimate of the function $f(.)$ is interpreted as a combined effect of scale, technological, compositional, and other possible effects, since technological change potentially depends on income and emissions. However, pairwise differencing identifies the pure effect of income growth on emissions. Therefore, the estimated income effects are interpreted as scale effects, and time effects reflect a composite effect, which is argued to be dominated by technological change and compositional effects in the theoretical EKC literature (see, for example, Taylor and Copeland (2004)). In theory, properly identified scale effects must be increasing in GDP per capita, which puts a check

[22] See Wagner (2008) for a discussion of these problems. Pairwise differencing is a possibilty to account for these problems.

on our results. An inverted U-shaped relation in total emissions can only arise, if the time effects are negative and strong enough to offset the scale effects.

Non-parametric estimations may suffer from end-of-sample biases, which may drive out-of-sample predictions. Therefore, we prefer a parametric approach in making in-sample estimations of equation (3.2) to estimate the income effects prior to the projections. However, compared to parametric estimation strategies, non-parametric ones have the advantage of imposing far less structure on the income effects. For this reason, we base our pair selection procedure, the GoF prior, on non-parametric estimations. As the non-parametric estimation strategy, we follow Schienle (2011) where the proposed non-parametric estimator applies to the cases with multiple nonstationary covariates.[23] The chosen pairs are intuitive, such as Western Offshoots with Western Europe, Eastern Europe with Former USSR, India with Latin America. For these six regions, every individual best pair is also the best pair of the other pair. However, this should not necessarily be the case for every region. Indeed, Latin America stands as the best pair also for China and Africa, and India is the best pair also for Other Asia.

3.3.2. Out-of-sample Extrapolation

In this section we briefly describe our extrapolation procedure. We start with extrapolating the income and time effects of each region. Next, the total effects of each region is obtained as the sum of the forecasted income and time effects. Finally, the extrapolations of the global income and total effects are the sum of the regional effects weighted by population. Global time effects are obtained as residuals. Before describing the extrapolation method of the individual series, we first formalize the aggregation process. In order to save on notation, in formulizing the aggregation process, we denote the levels of the series with the same notation which we used for the logarithm of the series in the previous subsection. Firstly, in line with the IPAT identity, the total effect for each region is given by

[23]The generalized smooth backfitting (G-SBF) estimator suggested by Schienle (2011), which applies to the cases with multiple nonstationary covariates, reduces to classical smooth backfitting (C-SBF) estimator (Mammen et al., 1999), under the assumption that the two dimensional nonstationarity of the paired GDP pc. series is as nonstationary as in the univariate GDP pc. series. Following this result, we implement classical SBF as explained inNielsen and Sperlich (2005).

$$\hat{y}_{i,T+h|T} = \hat{f}_{i,T+h|T}\hat{\lambda}_{i,T+h|T},$$

where $\hat{f}_{i,T+h|T}$ is the predicted levels of income effects for region i, h year ahead of the final sample year T, based on the information available at time T. Forecasts of regional time effects are denoted with $\hat{\lambda}_{i,T+h}$. The forecasts of total effects, $\hat{y}_{i,T+h|T}$, reflects the regional per capita impact. Forecasted levels of global total effects can be calculated as the population weighted averages of the forecasts of regional total effects as follows:

$$\hat{y}_{w,T+h|T} = \sum_{i=1}^{N} \frac{\hat{p}_{i,T+h|T}}{\hat{p}_{w,T+h|T}} \hat{y}_{i,T+h|T} \,,$$

where the subscript w indicates the series at the global level, p stands for population, and N is number of regions. In the same manner, global income effects can be calculated as the population weighted averages of the forecasts of regional income effects as follows:

$$\hat{f}_{w,T+h|T} = \sum_{i=1}^{N} \frac{\hat{p}_{i,T+h|T}}{\hat{p}_{w,T+h|T}} \hat{f}_{i,T+h|T} \,,$$

where $\hat{f}_{w,T+h|T}$ is the predicted levels of income effects for the whole world (w). Now, forecasts of the global time effects are denoted by $\hat{\lambda}_{w,T+h}$, and can be obtained as:[24]

$$\hat{\lambda}_{i,T+h|T} = \frac{\hat{y}_{i,T+h|T}}{\hat{f}_{i,T+h|T}}$$

Next, we describe how the individual series are extrapolated.

In order to perform the above procedure, first we need to extrapolate all the relevant series to the future. We make out-of-sample extrapolations with simple univariate ARIMA techniques, possibly augmented with linear trends. The forecasting model for each individual series is chosen from a pool of models. Candidate models are combinations of autoregressive and moving average terms up to order two, a linear or a quadratic time trend, and we also allow the series to be I(1). That is, candidate models are in the class of ARIMA models with deterministic trends. In choosing the forecasting model, we do not impose the order of integration a priori based on unit root tests. Instead, we allow the model selection criteria to choose the order of integration for each series as suggested by Chatfield (2002).

[24] Alternatively, one can work with series in logarithms, and the total effects can be derived by summing the logarithms of income and time effects.

In the main text, we present the results using BIC as the model selection criterion. This model selection criterion penalizes for increasing number of parameters, hence mitigates the over-fitting problem. In the main text, we present our results based on forecasts derived from models which are chosen according to their in-sample fit. Alternatively, one can divide the sample into an estimation period and a test period, and choose the model which predicts the test period better. This method also accounts for the over-fitting problem. In the Appendix, we present our results from model selection based on out-of-sample-fit evaluated with forecast mean squared errors (MSE). Our main results are robust to these considerations. Selected models for each individual series, their extrapolations, and some diagnostic tests are presented at the Appendix.

Taking logarithms of GDP and population series is a common application in econometric modeling for various purposes. In the context of forecasting, the goal of a logarithmic transformation might be to obtain a series with a relatively stationary variance. In their simulation, Lütkepohl and Xu (2012) show that a logarithmic transformation improves forecasts, only if the variance of the level variable is stationarized. On the other hand, forecasting in logarithms may result in dramatic distortions in forecast accuracy, if the level variable has already a stationary variance. Based on this finding, we allow the model selection process to choose whether a logarithmic transformation is required. Another issue in case of modeling the log-transformed variable is about transforming forecasts back into levels. Simply exponentiating the log-forecasts in order to obtain level-forecasts might not be optimal. Instead, following Granger and Newbold (1976), we apply the following transformation:

$$\hat{y}_{t+h|t} = \exp(\ln(\hat{y})_{t+h|t} + \frac{1}{2}\hat{\sigma}^2_{\ln(y)}),$$

where $\hat{\sigma}^2_{\ln(y)}$ is the mean squared error of log-forecasts.

3.4. Data and Descriptive Statistics

Our dataset is a balanced panel for all countries, covering the period between 1950 and 2010. CO_2 emission data consist of the sum of emissions from gas, liquid and solid fuels (based on consumption figures), and from gas flaring and cement production (see

Boden et al., 1995; Marland et al., 2009; Boden and Andres., 2013). For each type of fuel, data on annual CO_2 emissions result from three aspects: the amount of fuel consumed, the fraction of the fuel that becomes oxidized, and a factor for the carbon content of the fuel. The fuel types incorporated in the calculations are coal, other solid fuels, crude oil, petroleum products, and natural gas. Total energy use and emissions per country are corrected for exports and imports of fuels, as well as for stock changes, international marine bunkers, and non-energy use of fuels, such as chemical feedstock. The estimation of the amounts of CO_2 released through gas flaring are based on the UNSTAT database, supplemented by estimations from DOE/EIA. The estimations of the amounts of CO_2 released from cement manufacturing are based on figures indicating the quantity of manufactured cement, the average calcium oxide content per unit of cement and a factor to convert the calcium oxide content into CO_2 equivalents. Data on GDP and population are taken from Bolt and van Zanden (2013). All figures are expressed in 1990 International Geary-Khamis dollars, using purchasing power parities. Following Maddison (2009), we aggregate data on a country by country basis into nine regions: India, China, "Other Asia", western Europe, Eastern Europe, former USSR, "Western Offshoots", Africa, and Latin America. In contrast to the division into regions by the IPCC, we distinguish explicitly between Eastern Europe and Former USSR, divide the "old" OECD in Western Europe (old EU) and what we indicate as "Western Offshoots" (Australia, Canada, New Zealand, and the United States), while Japan, together with the countries of the Middle East are grouped under the name "Other Asia". Finally, we split the IPCC region ALM into Africa and Latin America. Figures 3.1 and 3.2 present our basic data.

Looking at our data on the distribution of GDP per capita (see Figure 3.1), Western Offshoots have by far the highest income per capita, whereas, in particular, India and Africa are on the lowest end of the scale. Clearly, the distribution has changed remarkably over time. At the beginning of our sample period, there were three clubs with Russia, Eastern Europe, and Latin America forming a rather stable middle-income group. Because of instability in these middle income regions as well as the remarkable growth for Other Asia and China since the 1990s, the set of middle-income countries currently contains five out of our nine regions.

Interestingly, both the distribution and development over time, the region-specific

Figure 3.1: GDP per Capita (Mln. US $ 1990)

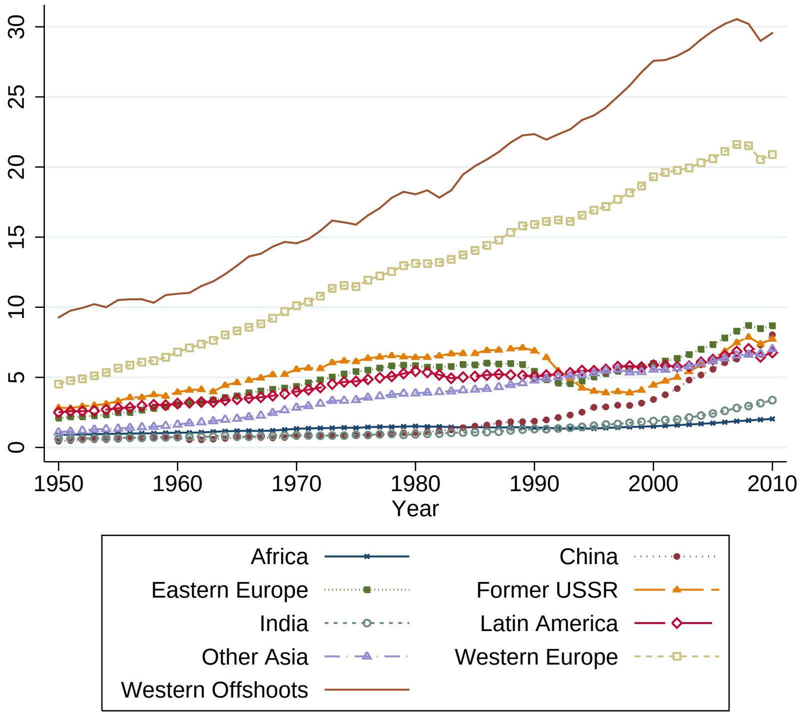
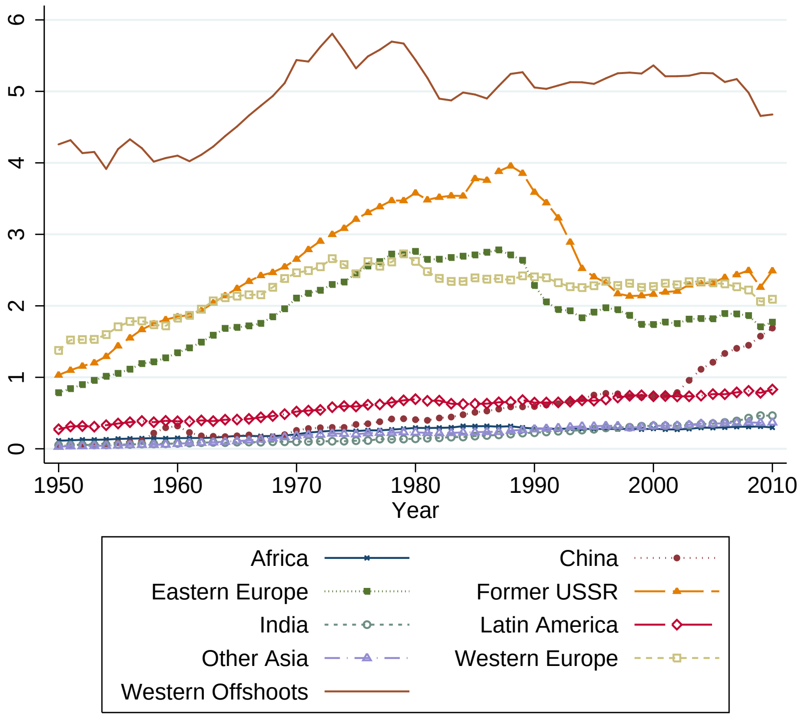


Figure 3.2: Carbon dioxide emission in tonnes per capita



per-capita CO_2 emissions are remarkably different (see Figure 3.2). The carbon intensity in the Western Offshoots has always been much higher than in any other region, followed by Former USSR, Western and Eastern Europe. Since these emissions reached a peak in Western Europe in the 1970s, carbon intensity there has remained more or less constant, whereas Former USSR and Eastern Europe have experienced a strong decline in emissions since the beginning of the 1990s. Most remarkable, however, is the recent, very high growth rate in China. China's growth in carbon intensity since 2001 is almost unprecedented. The only precursor in growth in per-capita carbon intensity since World War II, is the development in Western Offshoots during the 1960s. Indeed, China's per-capita carbon emissions have already reached the level of Eastern Europe of 2010.

Table 3.1: Descriptive statistics

	Units	Mean	Median	St. Dev.	Min.	Max.
Emission	Tons(mln)	534.976	336.746	477.245	18.171	2259.856
GDP	1990dollar(bln)	3293.401	1673.159	4503.200	183.017	29058.937
Population	mln	689.883	372.464	794.709	87.637	4138.919
Emission pc.	kg.	1483.931	699.753	1559.259	30.604	5806.490
GDP pc.	1990dollar	6118.600	4277.080	6369.410	448.022	30547.928
Emission pc. (log)		6.591	6.551	1.325	3.421	8.667
GDP pc. (log)		8.234	8.361	1.016	6.105	10.327

Note: Descriptive statistics are for the period 1950 - 2006. Total number of observation is 513.

Table 3.1 shows descriptive statistics of the data. Our data-set, aggregated over the regions, contains 9 regions and 61 years, resulting in 549 observations for all variables in our panel of CO_2 emissions.

3.5. Results

In this section, we present our regional and global emission projections. There are two patterns in the regional forecasts depending on the slope of the time effects. For the developed regions, Western Offshoots and Western Europe, the time effects are negatively sloped, possibly indicating that contribution of technological and compositional effects to CO_2 emissions are decreasing. For the other regions, the time effects are positively sloped. Here, we only present the results for China, representing the developing regions, and Western Offshoots, representing the developed regions. Interpretations of the results extends to other regions in these two groups.

In line with theoretical arguments in the EKC literature, in the analysis below, we interpret the time effects as a composite effect, reflecting not only the effect of technology but also industrial composition.

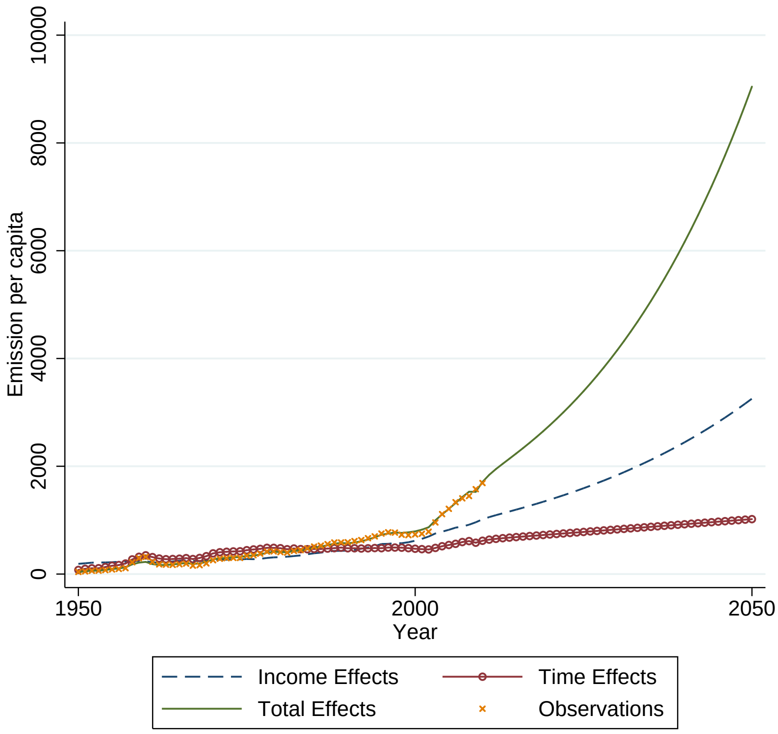
3.5.1. Is the developments in green technologies sufficient to reduce emissions at the regional and global level?

Figure 3.3 illustrates the in-sample estimates and the projections for China. Since the levels of the curves are not identified in the semi-parametric specifications, we normalize the curves per region in such a way that the average log-levels equals the corresponding sample average of the logarithm of CO_2 emissions per capita. In case of the income effect, we plot the $f(x_{it}, i)$ for the given region, i, as a function of time t, so that we actually plot the income effect using the income level at time t. Thus, moving from 1990 to 1991, the figures show the effect of the change in per-capita income between 1990 and 1991. Similarly, the time effect in the figure represents the estimated technological plus compositional effects for an additional year. Finally, the total effect just consists of the income effect plus the time effect at time t.

The results in Figure 3.3 illustrate that both income and time effects are increasing for China. That is, both the effect of the growing scale of the economy, and the combined effects of technological change and change in industrial composition are strong contributors to the emissions of China. In the corresponding future projections, this pattern is not likely to change for the period up to 2050, unless there is a structural change. This pattern is qualitatively the same for other developing regions.

The results for Western Offshoots are presented in Figure 3.4. Being in line with the theoretical arguments in the EKC literature, the income effect is increasing, while the time effect is decreasing. Therefore, the technological and compositional effect mitigates the increase in emissions due to a growing scale of the economy. Furthermore, it seems that there is a stabilization in emission per capita, which is likely to create a slow-down

Figure 3.3: In-sample Estimates and Extrapolations for CO_2Emissions Per Capita of China



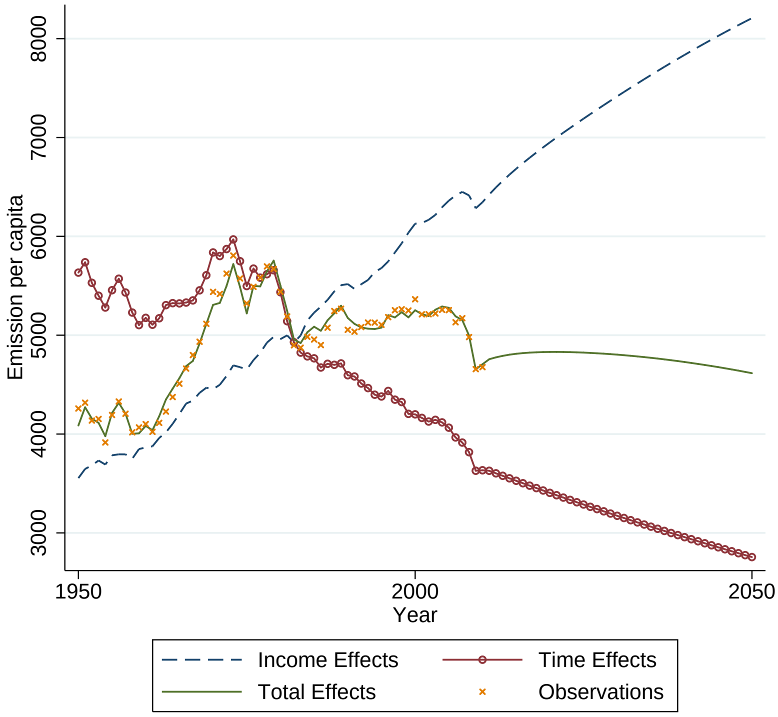
in emissions.

Our regional in-sample estimates and forecasts illustrate how pairwise differencing leads to results in-line with theory, by allowing for full-flexibility in the specification of the time effects. Thus, properly identified income effects seem to reflect pure scale effects, as is illustrated by the estimated income effects that are increasing.[25] The common inverted U-shaped and N-shaped findings in the EKC literature likely have to be interpreted as the combined effects of scale, technology, and industrial composition.

The pairwise differencing approach has some important advantages in forecasting. Firstly, the potentially nonlinear total effect is decomposed into its determinants, which are more likely to exhibit monotonic trends. Clearly, the estimated inverted U-shaped or N-shaped total effects in the EKC literature are not appropriate for forecasting purposes, since they often lead to forecasts with explosively increasing emissions, or zero emissions in the long-run. On the other hand, as illustrated in Figure 3.3 and 3.4, our findings indicate that the individual income and time effects are likely to have monotonic trends.

[25]Indeed, for all regions, our estimated income effects are increasing reflecting the emissions as a result of growing scale of the economy. These results are given in the appendix.

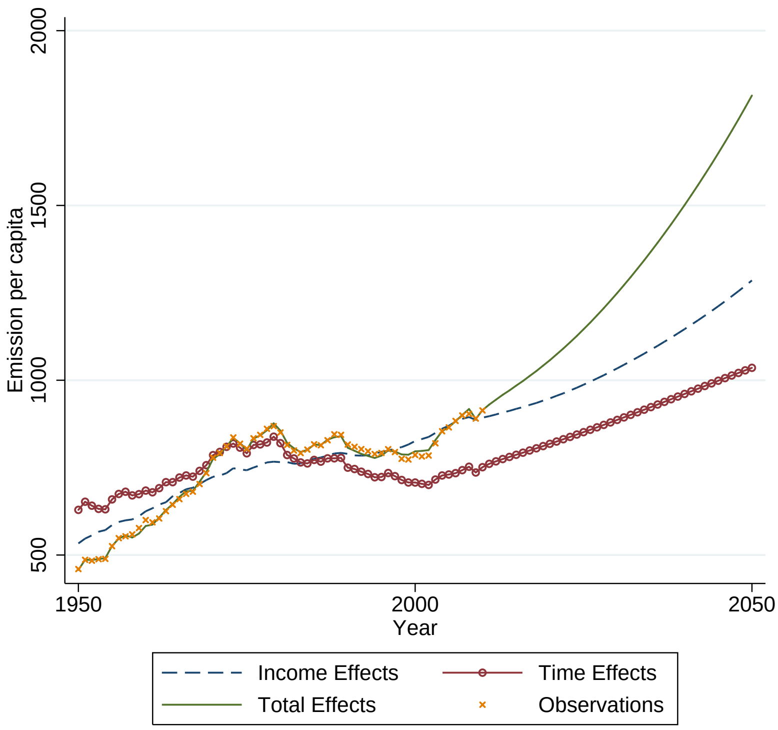
Figure 3.4: In-sample Estimates and Extrapolations for CO_2Emissions Per Capita of Western Off-shoots



Under the assumption of no trend break, extrapolating these monotonic individual trends are more convenient which prevents the counter intuitive results stemming from the use of the non-linear total effects. Moreover, Giacomini and Granger (2004) show that forecasting individual effects, instead of aggregate effects, can improve the forecasting performance. We achieve such a decomposition by decomposing the global effects into regional income and time effects.

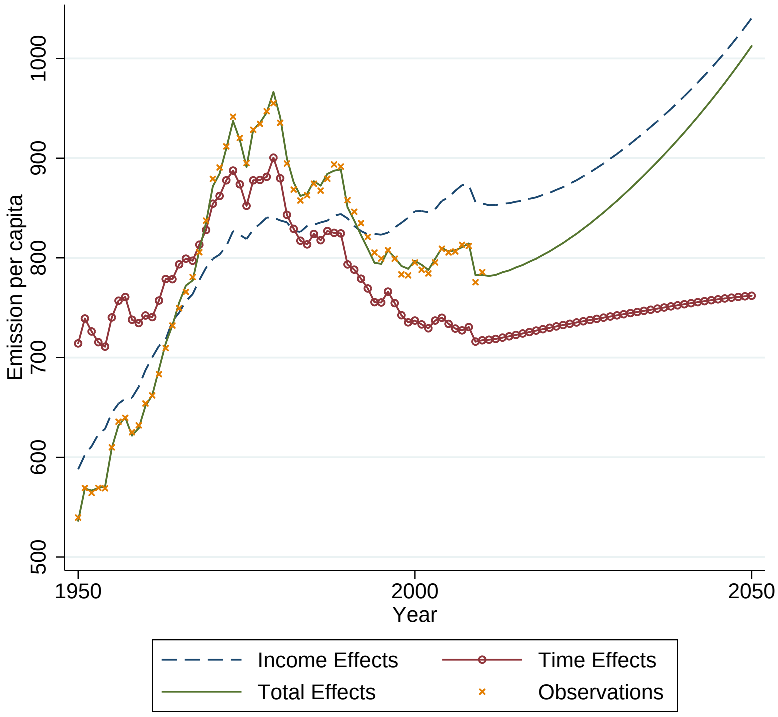
We make simple future extrapolations for the global emissions based on population-weighted averages for each region's best-fit estimates. To compute these global averages, we weight each of our region specific GoF estimates (after transforming our log estimates into levels) of the income and time effects with the region specific population levels. Similarly, we weight the future projections with projections of the population levels. The results in Figure 3.5 presents the development of our region specific findings. The pattern in the time effects are dominated by the developing world. That is, they are increasing and expected to increase in the future. However, this trend in time effects is not sufficient to compensate the increasing income effects. As a result, our projection indicates a sharp rise in global emissions.

Figure 3.5: In-sample Estimates and Extrapolations of Global CO_2 Emission Per Capita



In theory, the effect of industrial composition should constitute a less important role in the time effects at a global level, compared to its role at the regional level. The reason is that the change in sectoral composition, affecting the emission level, is hypothesized to be mostly driven by a shift of dirty industries from regions with strict environmental regulations to the regions with less strict regulations (Pollution Haven Hypothesis). Clearly, such an effect should cancel out at the global level, since it is a mere replacement of dirty industries. Another reason for the change in sectoral composition can be the directed technological change towards cleaner technologies, which can be considered as a technological effect rather than a compositional effect. Therefore, in theory, global time effects should constitute technological effects. Given this explanation, we can answer the question whether technological progress will be sufficient to create a slow-down in future emissions. Our global level forecasts illustrate that such a slow-down is not likely, indicating a pessimistic picture. Although, there is some mitigating effect of technological change in the developed regions, these are far from being sufficient to compensate the effect of a growing scale of the global economy.

Figure 3.6: In-sample Estimates and Extrapolations of Global CO_2 Emission Per Capita Excluding China



3.5.2. Is China the main threat in combating with global warming?

Our finding of a positive time effect for China reflects their recent switch to a coal-based energy input mix, as well as their strong industrial expansion, both of which have contributed to the recent upsurge in global carbon emissions. However, this trend has not co-evolved with a strong enough negative time effect in developed regions, such as Western Europe and Western Offshoots, in order to induce an overall reduction in global per-capita carbon emissions. In fact, the underlying regional trends in emission patterns make a reversal of the overall global trend quite unlikely, for the next decades at least.

These findings strengthen the concerns that high growth rates by China, which are expected to continue for the following decades, may constitute the main problem for the struggle against environmental degradation. In order to analyze if this is the case, we forecast global emissions by excluding China. Figure 3.6 illustrates the results. Although the recent strong growth in per-capita emissions in China certainly have contributed to the renewed upward overall trend (Figure 3.5), the same result is obtained when we exclude China from the sample. This suggests that the underlying current developments in the other regions have been such that the downward sloping time effect cannot compensate for the strong positive income effect. This remarkable result is at odds with the popular view that particularly China would be the most important threat to the policies that aim at stabilizing global carbon emissions.

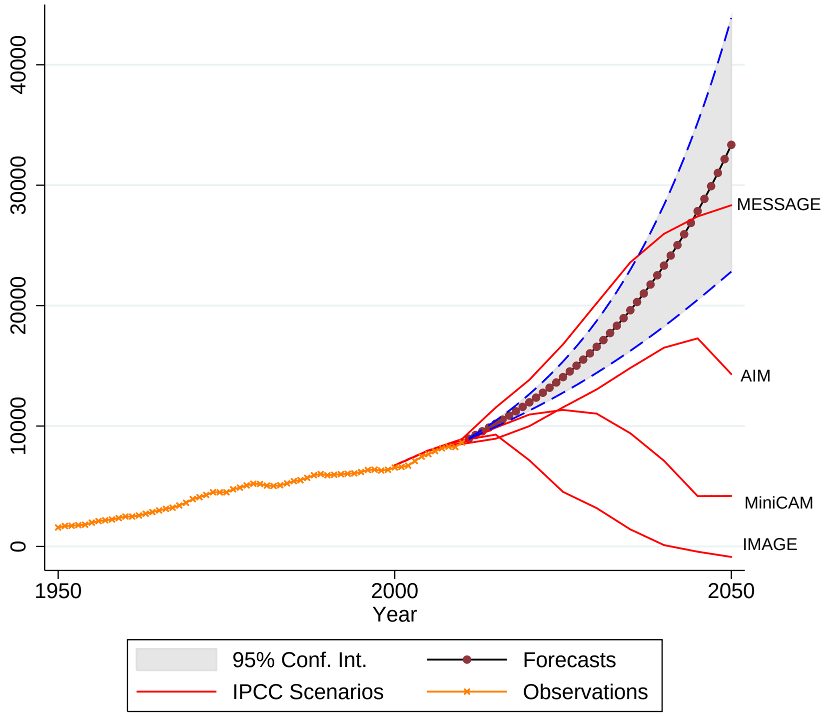
3.5.3. Comparison with IPCC Scenarios

In this section we compare our extrapolations with some representative IPCC scenarios. In Figure 3.7, four different IPCC scenarios are presented. These projections are called "Representative Concentration Pathways" (RCP) which are selected to represent the wide range of scenarios published in the literature at 2007 (see Moss et al. 2010 for a detailed discussion about the selection process). In Figure 3.7, each individual RCP is labeled by the name of its modelling team. The most optimist one among the four, IMAGE (RCP 2.6) (van Vuuren et al., 2007), represents the 10th percentile of the mitigation scenarios. MESSAGE (RCP 8.5) (Riahi et al., 2007), being a pessimist one, represents the 90th percentile of the range spanned by emission scenarios by IPCC. MiniCAM (RCP 4.5) (Clarke et al., 2007; Smith and Wigley, 2006; Wise et al., 2009) and AIM (RCP 6.0) (Fujino et al., 2006; Hijioka and Kainuma, 2008) stands between these two extreme scenarios.

In Figure 3.7, we also present the 95 percent confidence intervals based on our extrapolations. The confidence intervals includes the uncertainity of all the parameters in our model. Our baseline extrapolations supports the range spanned by IPCC scenarios between the 90th and 60th percentile. Except MESSAGE, other three scenarios predicts a peak until 2050, which seems unlikely according to our 95 percent confidence intervals constructed from our baseline extrapolations. Our confidence intervals also exclude the most pessimistic scenario, MESSAGE, at its earliar stage.

An important point is that the presented IPCC projections are based on no-policy scenarios. Our results indicate that a peak in the emissions up to 2050 is not likely unless there is a trend break such as strong policy interventions implemented in the future. That is, the current pace of the changes in income-related emissions and in potentially mitigating factors are not likely to create a slow-down in our forecast period.

Figure 3.7: Global CO_2 Emissions and IPCC Scenarios



3.6. Conclusion

The pairwise differencing approach is a flexible way to disentangle the scale effects from other factors. Our results show that such a decomposition can reveal some powerful insights about the underlying trends driving the global CO_2 emissions, and provides a convenient tool to make future extrapolations. Our results reveal a pattern consistent with the role of scale effects in the growth and environment literature that the income effects are positive for any region. That is, GDP per capita and the corresponding amount of emissions are always positively linked. This shows that the pairwise differencing approach is able to identify the scale effects properly.

While we find a strong income effect leading to a sharp rise in global emissions, the global time effects, dominated by the trends in the developed regions, play a mitigating role, but far from compensating the strong income effects. As a result, our extrapolations do not even imply a slow down in global CO_2 emissions up to 2050. We further present an analysis regarding the recent role of China in the global emission pathways. We find that the pessimistic patterns revealed by our forecasts is not driven by China alone, but another important problem is the weakness of the time effects in the developed regions to create a global U-turn.

As a final note, there are some points in our analysis that are open to be improved by future research. Firstly, our analysis fully relies on historical data, including the projections of GDP and poulation series, by which we aim to avoid subjective uncertainty. However, expert judgements about the future evolution of these series can easily be incorporated in our analysis in order to produce more scenario-based forecasts. For example, our future forecasts of population for China implies a downturn around 2030, which can be a reasonable forecast due to the one-child policy of China. However, if one believes only a slow-down, but not a downturn as a possible scenario, this can be easily incorporated as a constraint in the forecasting model. Furthermore, since we are able to decompose the income and time effects, such scenarios can also include expert views on these series.

A second point is that our pairwise differencing estimations rely on the GoF prior in order to match regions which have similar time effects. In our view, a matching strategy based on an expert judgment can also be legitimate. Indeed, the matching by GoF prior produce intuitive results, such as matching Western Europe with Western Offshoots, or Eastern Europe with Former USSR. Any estimation in the EKC literature applies such prior beliefs, like imposing homogeneous income and time effects in the panel estimations. However, using expert judgment can be infeasible when there are many cross-sectional units, and the GoF prior becomes a necessity. Using another matching process, possibly relying on observed data relevant to the time effects, can be a possible future research avenue.

3.A. Appendix

3.A.1. Model selection based on out-of-sample performance

In this section, we present our results from using the alternative model selection process where we decompose our data into an estimation and a test sample. We estimate the candidate models with the sample up to 1995. We make the model selection based on their performance in predicting the observed data from 1996 to 2010. We assess the fit with forecast mean squared errors of the candidate models.

Figure 3.8: Extrapolations for China with Out-of-sample Fit

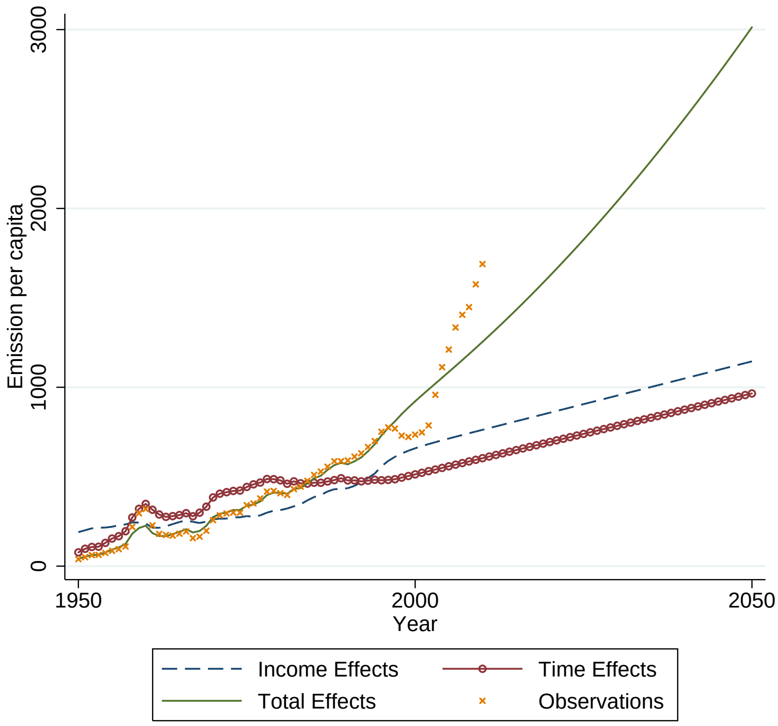


Figure 3.9: Extrapolations for Western Offshoots with Out-of-sample Fit

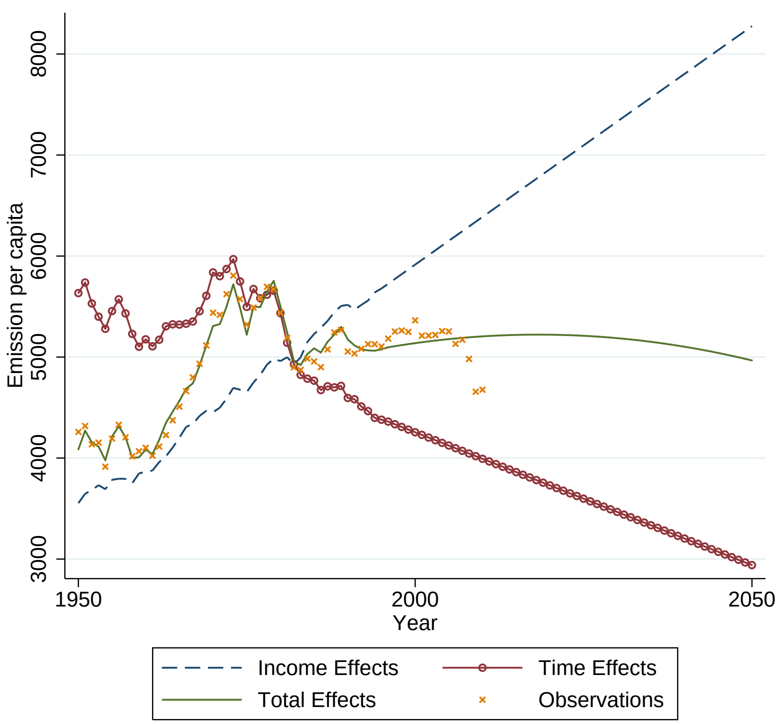


Figure 3.10: Extrapolations of Global Effects with Out-of-sample Fit

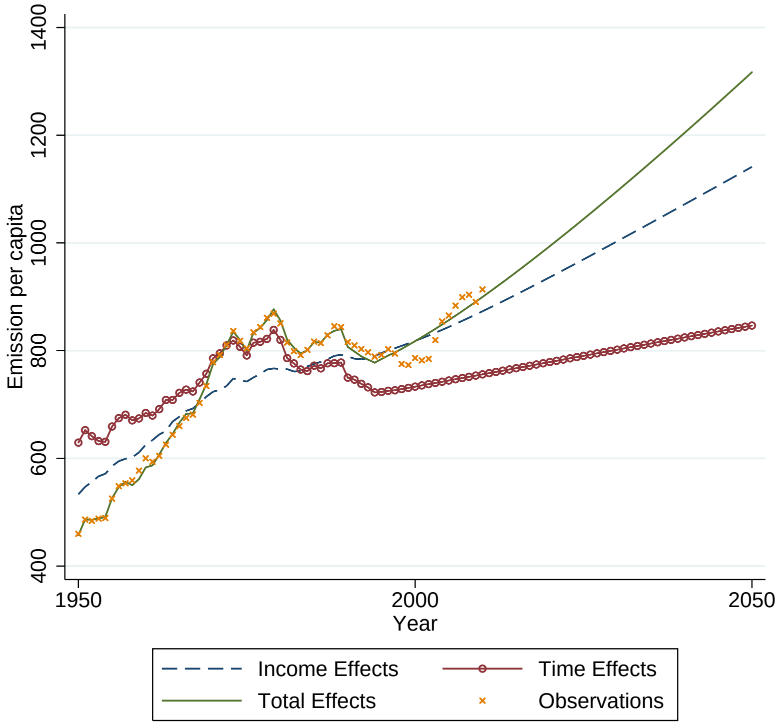


Figure 3.8 and 3.9 present the extrapolation results from using the out-of-sample fit method for model selection. The results are qualitatively similar. The only remarkable difference with our baseline extrapolations is that the increase in income effects of China is now weaker. Similarly, the extrapolations of global effects are again similar to our baseline extrapolations as presented in Figure 3.10.

A comparison with the IPCC scenarios is presented in Figure 3.11. The general trend is similar to our baseline estimations, only with the difference that AIM slightly stays in our confidence intervals until it peaks around 2040.

3.A.2. Estimation Tables

Table 3.2: Pairwise Diffirenced ENNLS Estimations Africa

	(1)	(2)	(3)	(4)	(5)
GDP pc.	0.330*	-18.601	593.753		
	(0.183)	(11.801)	(374.078)		
GDP pc.^{2}		1.296	-82.888	40.236	
		(0.808)	(51.408)	(26.035)	
GDP pc.^{3}			3.857	-7.486	3.630
			(2.355)	(4.766)	(2.423)
GDP pc.^{4}				0.392	-0.759
				(0.245)	(0.499)
GDP pc.^{5}					0.042
					(0.027)
Pair-GDP pc.	887.479***	868.708***	1834.764***	1808.438***	1781.608***
	(321.327)	(316.992)	(664.891)	(647.254)	(631.060)
Pair-GDP pc.^{2}	-58.732***	-58.238***	-121.014***	-119.290***	-117.533***
	(21.055)	(20.763)	(43.280)	(42.129)	(41.072)
Pair-GDP pc.^{3}	1.296***	1.300***	2.660***	2.622***	2.584***
	(0.460)	(0.454)	(0.939)	(0.914)	(0.891)
Adjusted R^2	0.329	0.347	0.368	0.368	0.369
AIC	-155.6	-156.4	-157.3	-157.3	-157.3
BIC	-147.3	-146.0	-144.8	-144.9	-144.9
Observations	59	59	59	59	59

Table 3.3: Pairwise Diffirenced ENNLS Estimations China

	(1)	(2)	(3)	(4)	(5)
GDP pc.	0.580***	2.727*	10.205	-342.021*	
	(0.096)	(1.461)	(16.912)	(179.381)	
GDP pc.^{2}		-0.146	-1.135	69.209*	-21.674*
		(0.098)	(2.224)	(35.795)	(12.387)
GDP pc.^{3}			0.044	-6.174*	5.855*
			(0.097)	(3.165)	(3.275)
GDP pc.^{4}				0.205*	-0.588*
				(0.105)	(0.324)
GDP pc.^{5}					0.021*
					(0.011)
Pair-GDP pc.	-3411.581***	-4661.428***	-4451.421***	-4937.352***	-4856.458***
	(810.560)	(1141.808)	(1226.812)	(1209.169)	(1206.476)
Pair-GDP pc.^{2}	221.896***	304.755***	290.903***	322.350***	317.077***
	(53.185)	(75.267)	(80.824)	(79.610)	(79.445)
Pair-GDP pc.^{3}	-4.810***	-6.641***	-6.336***	-7.014***	-6.900***
	(1.163)	(1.654)	(1.775)	(1.747)	(1.744)
Pair-GDP pc.^{4}					
Pair-GDP pc.^{5}					
Adjusted R^2	0.908	0.909	0.907	0.912	0.912
AIC	-32.1	-31.8	-29.7	-32.1	-31.8
BIC	-23.8	-21.4	-17.2	-17.6	-17.3
Observations	59	59	59	59	59

Standard errors in parentheses

Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Pairwise Diffirenced ENNLS Estimations Eastern Europe

	(1)	(2)	(3)	(4)	(5)
GDP pc.	0.193***	-0.103	-23.622		-101.314
	(0.037)	(0.910)	(29.692)		(257.954)
GDP pc.^{2}		0.018	2.823	-1.393	
		(0.054)	(3.541)	(1.784)	
GDP pc.^{3}			-0.111	0.223	2.784
			(0.141)	(0.283)	(7.367)
GDP pc.^{4}				-0.010	-0.326
				(0.013)	(0.879)
GDP pc.^{5}					0.011
					(0.031)
Pair-GDP pc.	0.256***	0.253***	0.246***	0.246***	0.248***
	(0.049)	(0.050)	(0.050)	(0.051)	(0.051)
Pair-GDP pc.^{2}					
Pair-GDP pc.^{3}					
Pair-GDP pc.^{4}					
Pair-GDP pc.^{5}					
Adjusted R^2	0.322	0.311	0.303	0.303	0.293
AIC	-169.3	-167.4	-166.0	-166.0	-166.1
BIC	-163.1	-159.1	-155.7	-155.6	-155.8
Observations	59	59	59	59	59

Standard errors in parentheses

Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Pairwise Diffirenced ENNLS Estimations Former USSR

	(1)	(2)	(3)	(4)	(5)
GDP pc.	0.256***	0.924	106.317		
	(0.049)	(1.855)	(83.290)		
GDP pc.^{2}		-0.039	-12.513	6.379	
		(0.109)	(9.852)	(4.937)	
GDP pc.^{3}			0.492	-0.999	0.514
			(0.388)	(0.778)	(0.390)
GDP pc.^{4}				0.044	-0.090
				(0.034)	(0.069)
GDP pc.^{5}					0.004
					(0.003)
Pair-GDP pc.	0.193***	0.196***	0.213***	0.214***	0.214***
	(0.037)	(0.039)	(0.040)	(0.040)	(0.040)
Pair-GDP pc.^{2}					
Pair-GDP pc.^{3}					
Pair-GDP pc.^{4}					
Pair-GDP pc.^{5}					
Adjusted R^2	0.322	0.311	0.327	0.327	0.328
AIC	-169.3	-167.5	-169.0	-167.0	-167.0
BIC	-163.1	-159.1	-160.7	-156.6	-156.6
Observations	59	59	59	59	59

Standard errors in parentheses

Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Pairwise Diffirenced ENNLS Estimations India

	(1)	(2)	(3)	(4)	(5)
GDP pc.	1.348***	10.433***	72.286***	-1139.634***	
	(0.079)	(0.992)	(21.299)	(323.338)	
GDP pc.^{2}		-0.642***	-9.097***	243.308***	-74.074***
		(0.067)	(2.911)	(67.340)	(22.835)
GDP pc.^{3}			0.385***	-22.938***	21.183***
			(0.133)	(6.225)	(6.323)
GDP pc.^{4}				0.807***	-2.255***
				(0.215)	(0.656)
GDP pc.^{5}					0.085***
					(0.024)
Pair-GDP pc.					
Pair-GDP pc.^{2}					
Pair-GDP pc.^{3}	2.113***				
	(0.544)				
Pair-GDP pc.^{4}	-0.211***				
	(0.054)				
Pair-GDP pc.^{5}	0.006***				
	(0.001)				
Adjusted R^2	0.955	0.970	0.974	0.979	0.979
AIC	-129.7	-152.5	-159.0	-172.3	-172.2
BIC	-121.4	-144.1	-148.6	-161.9	-161.8
Observations	59	59	59	59	59

Standard errors in parentheses

Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Pairwise Diffirenced ENNLS Estimations Latin America

	(1)	(2)	(3)	(4)	(5)
GDP pc.	0.973***	9.449**	60.282		
	(0.110)	(4.551)	(123.162)		
GDP pc.^{2}		-0.507*	-6.680	4.133	
		(0.272)	(14.888)	(7.308)	
GDP pc.^{3}			0.250	-0.612	0.376
			(0.600)	(1.176)	(0.578)
GDP pc.^{4}				0.026	-0.063
				(0.053)	(0.105)
GDP pc.^{5}					0.003
					(0.005)
Pair-GDP pc.					
Pair-GDP pc.^{2}	10.104***				
	(2.877)				
Pair-GDP pc.^{3}	-0.911***	0.727***	0.744***	0.744***	0.744***
	(0.269)	(0.177)	(0.186)	(0.186)	(0.186)
Pair-GDP pc.^{4}	0.023***	-0.074***	-0.076***	-0.076***	-0.076***
	(0.007)	(0.019)	(0.020)	(0.020)	(0.020)
Pair-GDP pc.^{5}		0.002***	0.002***	0.002***	0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
Adjusted R^2	0.974	0.975	0.975	0.975	0.975
AIC	-161.7	-163.6	-161.9	-161.8	-161.8
BIC	-153.4	-153.2	-149.4	-149.4	-149.4
Observations	59	59	59	59	59

Standard errors in parentheses

Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Pairwise Diffirenced ENNLS Estimations Other Asia

	(1)	(2)	(3)	(4)	(5)
GDP pc.	1.469***	17.111***	16.943		-508.560***
	(0.104)	(1.413)	(37.468)		(165.146)
GDP pc.^{2}		-1.003***	-0.979	2.685	
		(0.090)	(4.832)	(2.341)	
GDP pc.^{3}			-0.001	-0.349	16.872***
			(0.208)	(0.402)	(5.368)
GDP pc.^{4}				0.012	-2.158***
				(0.019)	(0.683)
GDP pc.^{5}					0.082***
					(0.026)
Pair-GDP pc.	24.064***				
	(4.103)				
Pair-GDP pc.^{2}	-0.797***				
	(0.142)				
Pair-GDP pc.^{3}		1.178***	1.172***	1.181***	1.030***
		(0.192)	(0.200)	(0.198)	(0.190)
Pair-GDP pc.^{4}		-0.122***	-0.121***	-0.122***	-0.107***
		(0.020)	(0.021)	(0.021)	(0.020)
Pair-GDP pc.^{5}		0.003***	0.003***	0.003***	0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
Adjusted R^2	0.854	0.952	0.950	0.951	0.957
AIC	-92.6	-159.1	-156.8	-157.1	-164.4
BIC	-84.2	-148.7	-144.3	-144.6	-149.9
Observations	59	59	59	59	59

Standard errors in parentheses

Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: Pairwise Diffirenced ENNLS Estimations Western Europe

	(1)	(2)	(3)	(4)	(5)
GDP pc.	0.831***	7.741***	50.799***		169.069
	(0.058)	(1.800)	(13.809)		(103.354)
GDP pc.^{2}		-0.396***	-5.302***	2.920***	
		(0.103)	(1.480)	(0.768)	
GDP pc.^{3}			0.186***	-0.405***	-3.739
			(0.053)	(0.110)	(2.424)
GDP pc.^{4}				0.016***	0.394
				(0.004)	(0.262)
GDP pc.^{5}					-0.012
					(0.008)
Pair-GDP pc.	0.958***	13.233***	0.499***	0.504***	0.477**
	(0.071)	(3.602)	(0.181)	(0.181)	(0.180)
Pair-GDP pc.^{2}		-0.385***			
		(0.110)			
Pair-GDP pc.^{3}					
Pair-GDP pc.^{4}					
Pair-GDP pc.^{5}					
Adjusted R^2	0.799	0.831	0.832	0.831	0.835
AIC	-243.5	-253.0	-253.8	-253.4	-255.8
BIC	-237.2	-242.6	-243.4	-243.0	-245.4
Observations	59	59	59	59	59

Standard errors in parentheses

Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: Pairwise Diffirenced ENNLS Estimations Western Offshoots

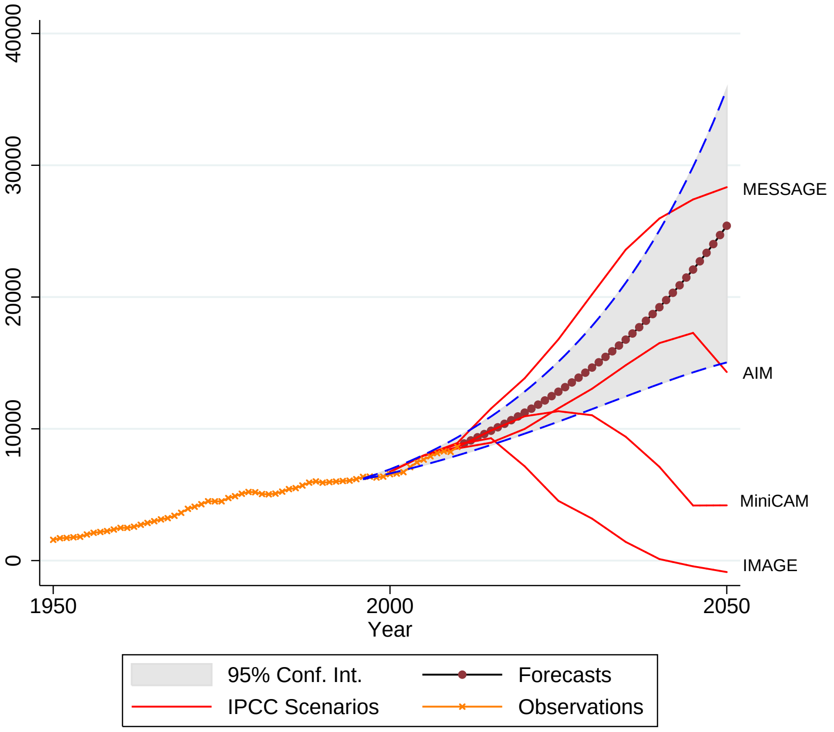
	(1)	(2)	(3)	(4)	(5)
GDP pc.	0.500***	7.913***	29.122		
	(0.181)	(2.079)	(50.940)		
GDP pc.^{2}		-0.385***	-2.518	1.944	
		(0.110)	(5.116)	(2.668)	
GDP pc.^{3}			0.071	-0.232	0.430*
			(0.171)	(0.357)	(0.246)
GDP pc.^{4}				0.008	-0.067*
				(0.013)	(0.039)
GDP pc.^{5}					0.003*
					(0.002)
Pair-GDP pc.	150.103***	13.214***	14.218***	14.186***	
	(41.799)	(3.229)	(4.112)	(4.056)	
Pair-GDP pc.^{2}	-9.118***	-0.396***	-0.428***	-0.427***	
	(2.572)	(0.104)	(0.131)	(0.129)	
Pair-GDP pc.^{3}	0.185***				0.485**
	(0.053)				(0.225)
Pair-GDP pc.^{4}					-0.045**
					(0.021)
Pair-GDP pc.^{5}					0.001**
					(0.001)
Adjusted R^2	0.832	0.831	0.829	0.829	0.833
AIC	-255.7	-253.0	-253.2	-253.2	-254.5
BIC	-247.4	-242.6	-242.9	-242.8	-242.0
Observations	59	59	59	59	59

Standard errors in parentheses

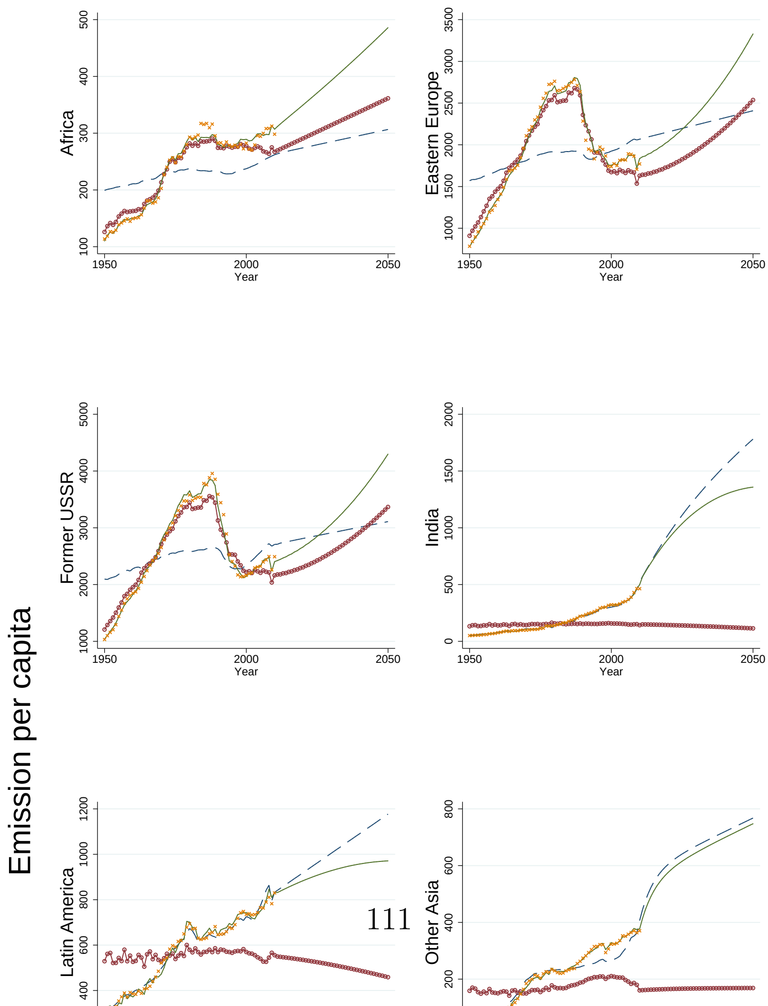
Note:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.11: Global CO_2 Emissions and IPCC Scenarios with Out-of-sample Fit

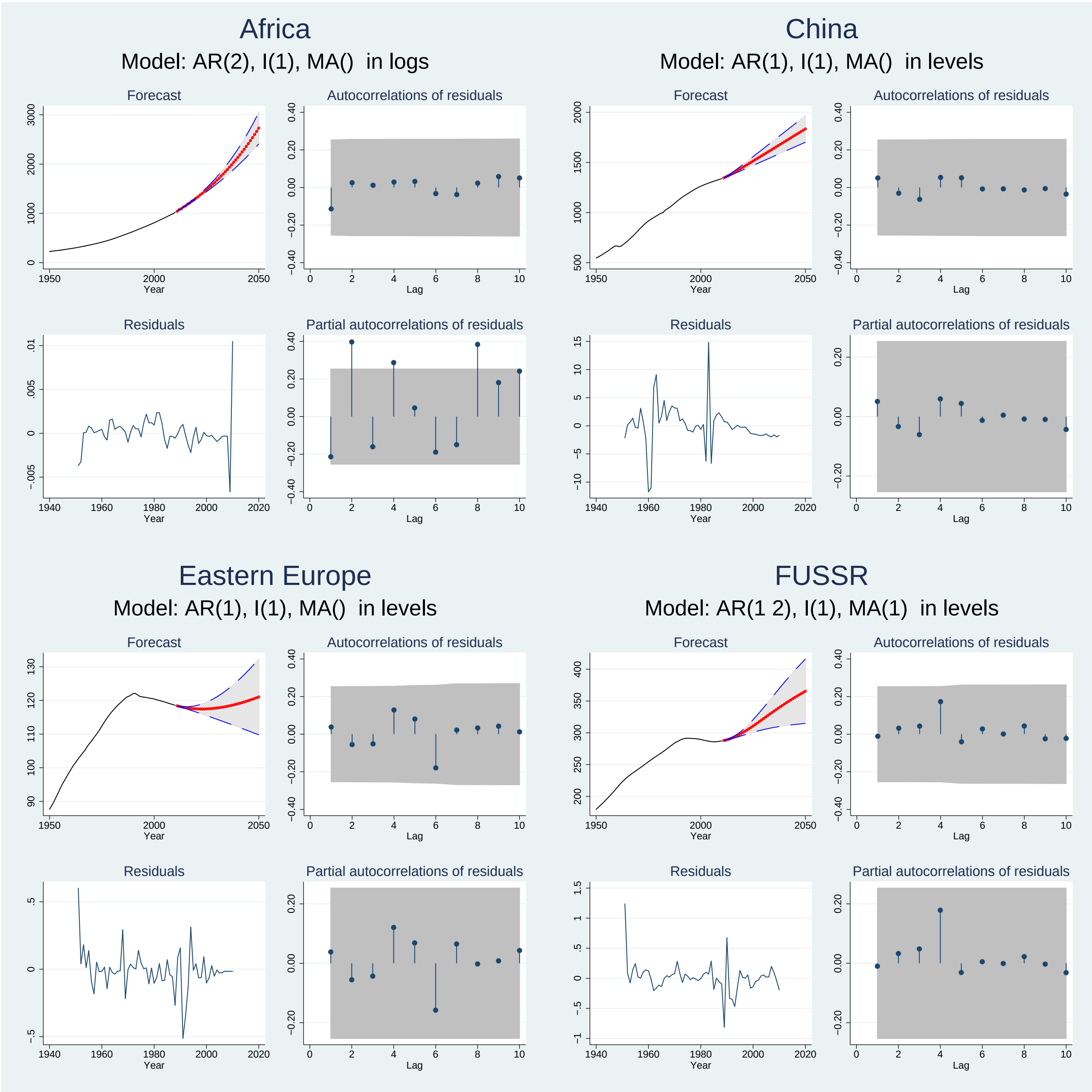


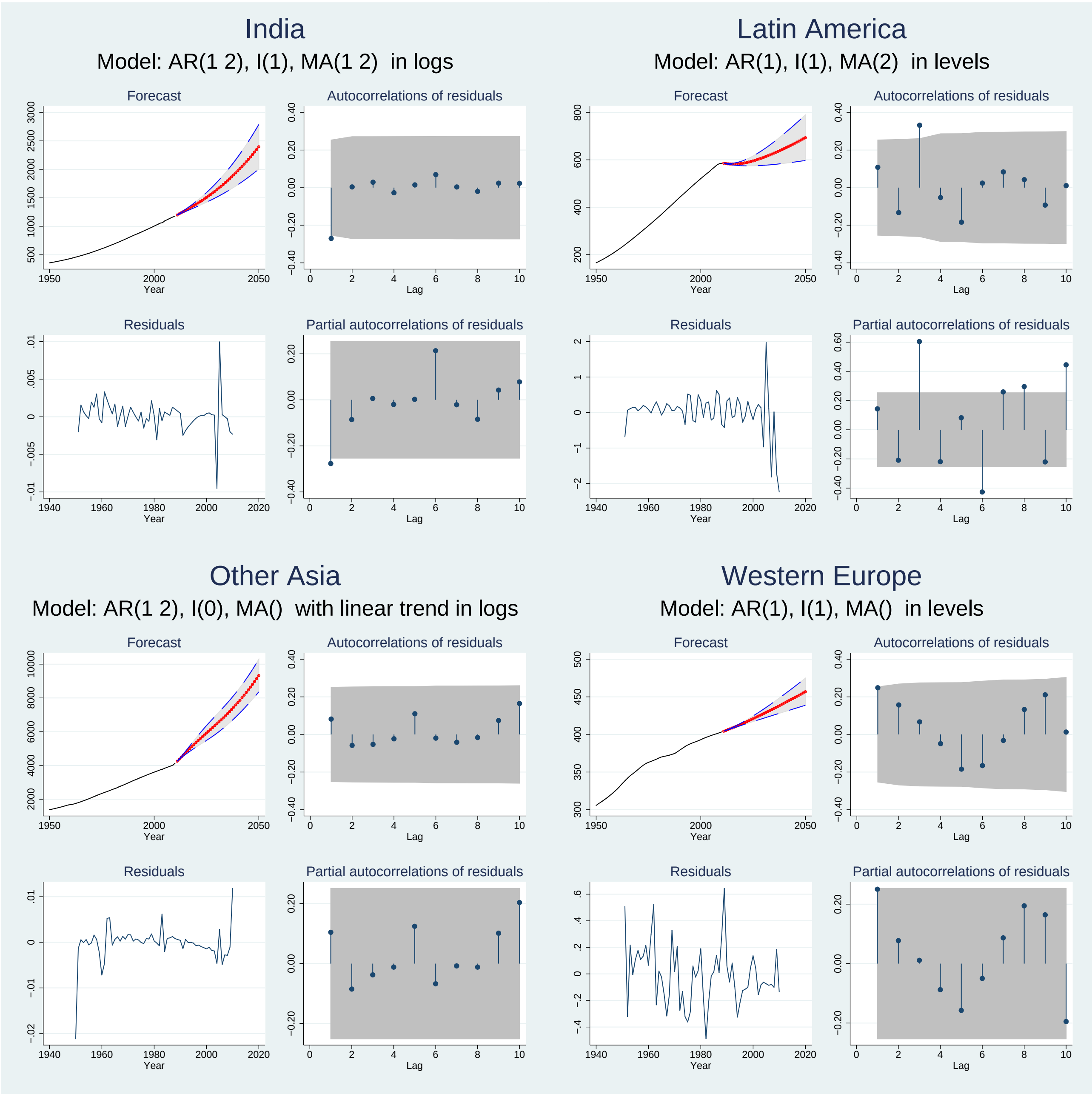
3.A.3. Extrapolations for Other Individual Regions

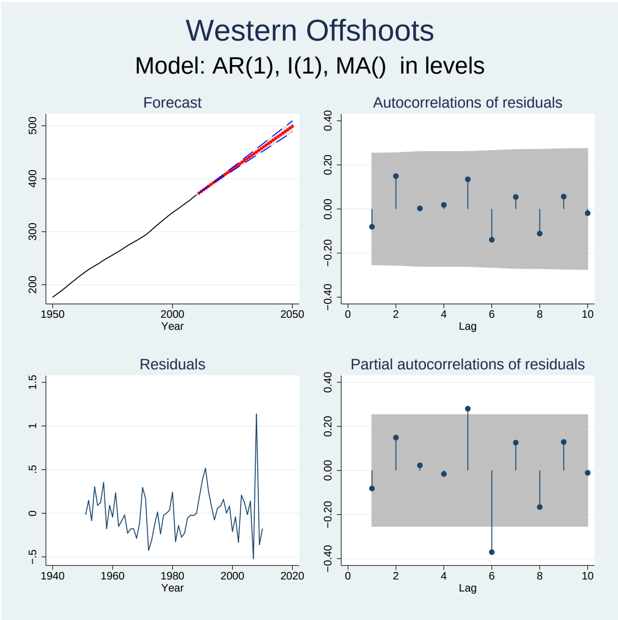


3.A.4. Extrapolations of Individual Series and Some Diagnostic Tests

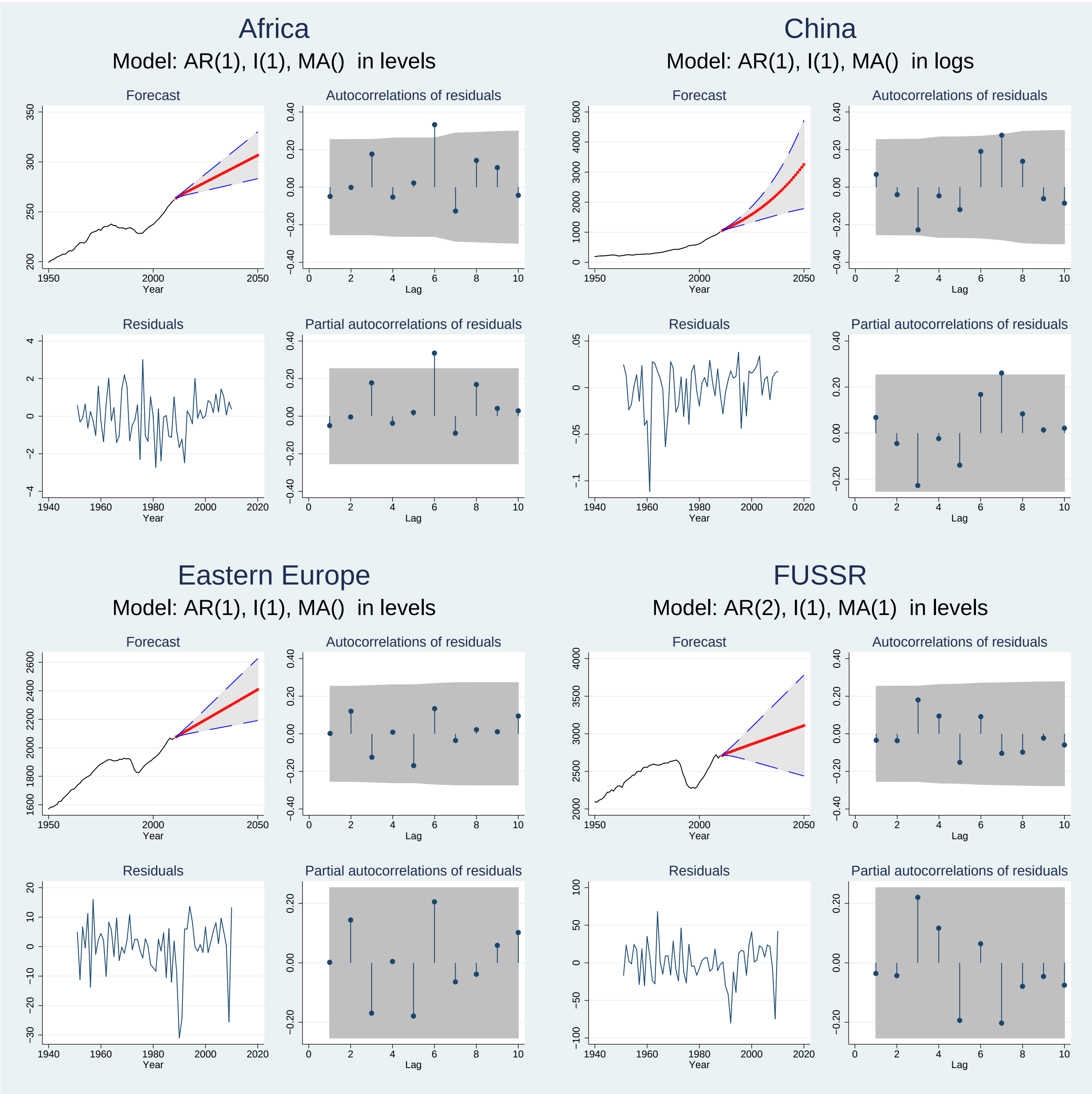
Population Series

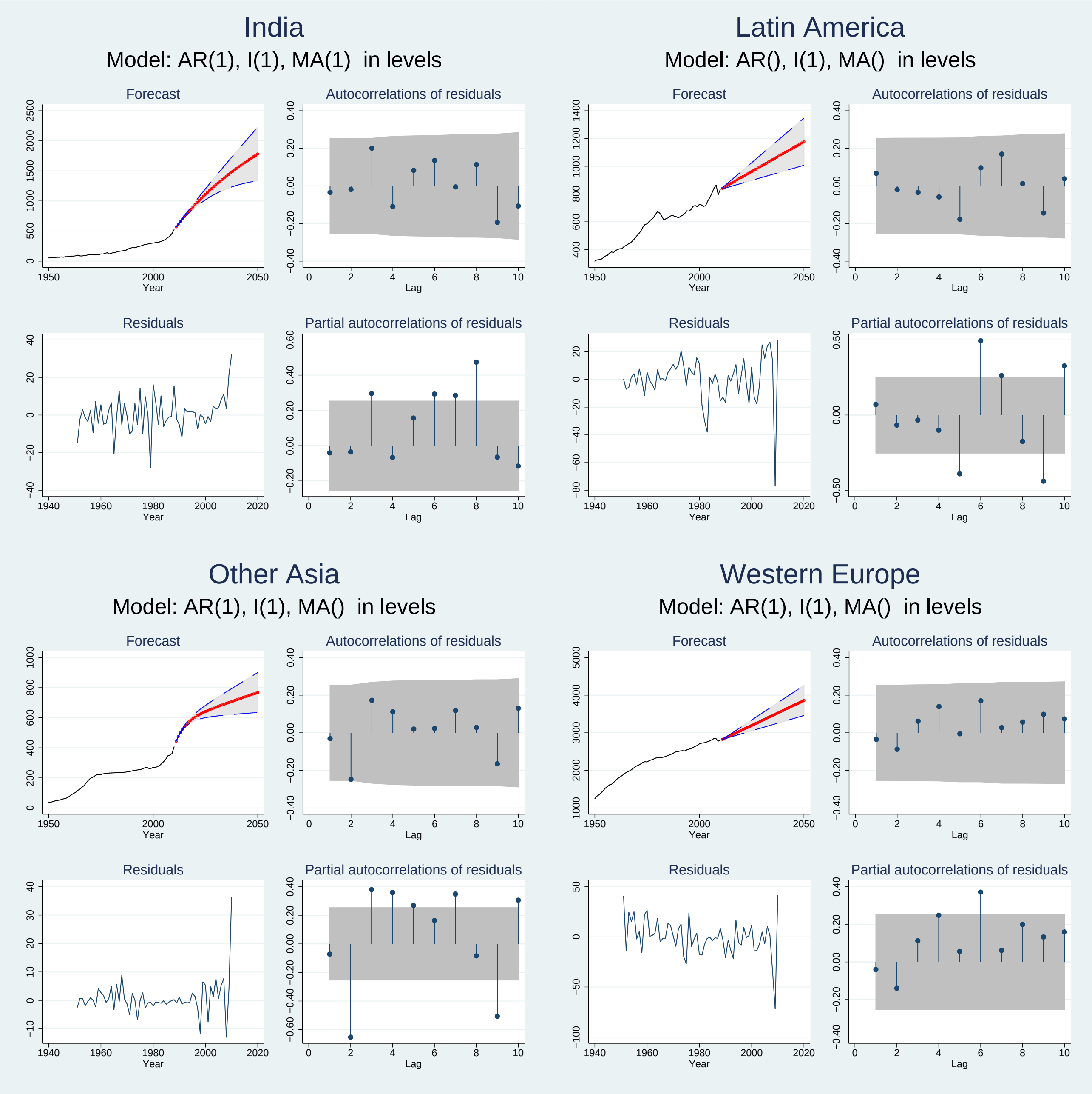


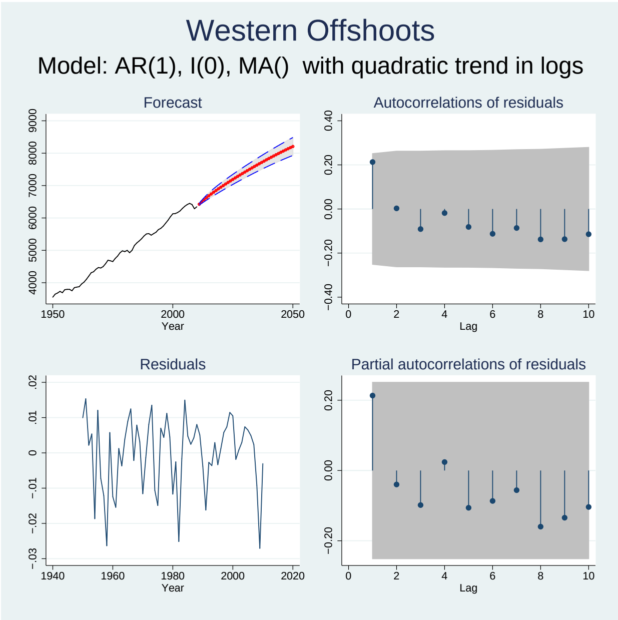

India
Model: AR(1 2), I(1), MA(1 2) in logs
Forecast
Autocorrelations of residuals
Residuals
Partial autocorrelations of residuals
Latin America
Model: AR(1), I(1), MA(2) in levels
Forecast
Autocorrelations of residuals
Residuals
Partial autocorrelations of residuals
Other Asia
Model: AR(1 2), I(0), MA() with linear trend in logs
Forecast
Autocorrelations of residuals
Residuals
Partial autocorrelations of residuals
Western Europe
Model: AR(1), I(1), MA() in levels
Forecast
Autocorrelations of residuals
Residuals
Partial autocorrelations of residuals
Year
Lag


Western Offshoots
Model: AR(1), I(1), MA() in levels
Forecast
Autocorrelations of residuals
Residuals
Partial autocorrelations of residuals
Year
Lag
Year
Lag

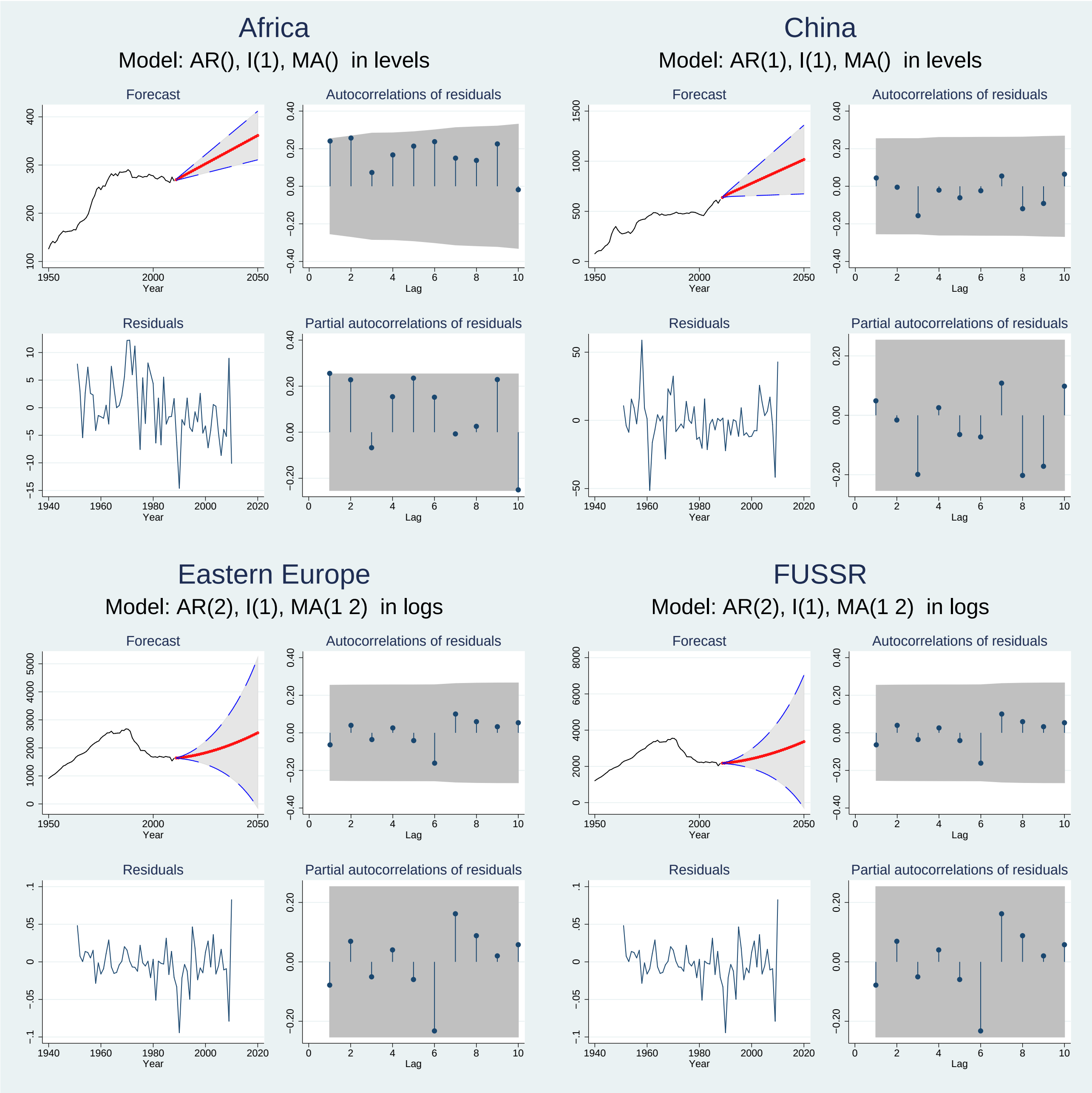
Income Effects

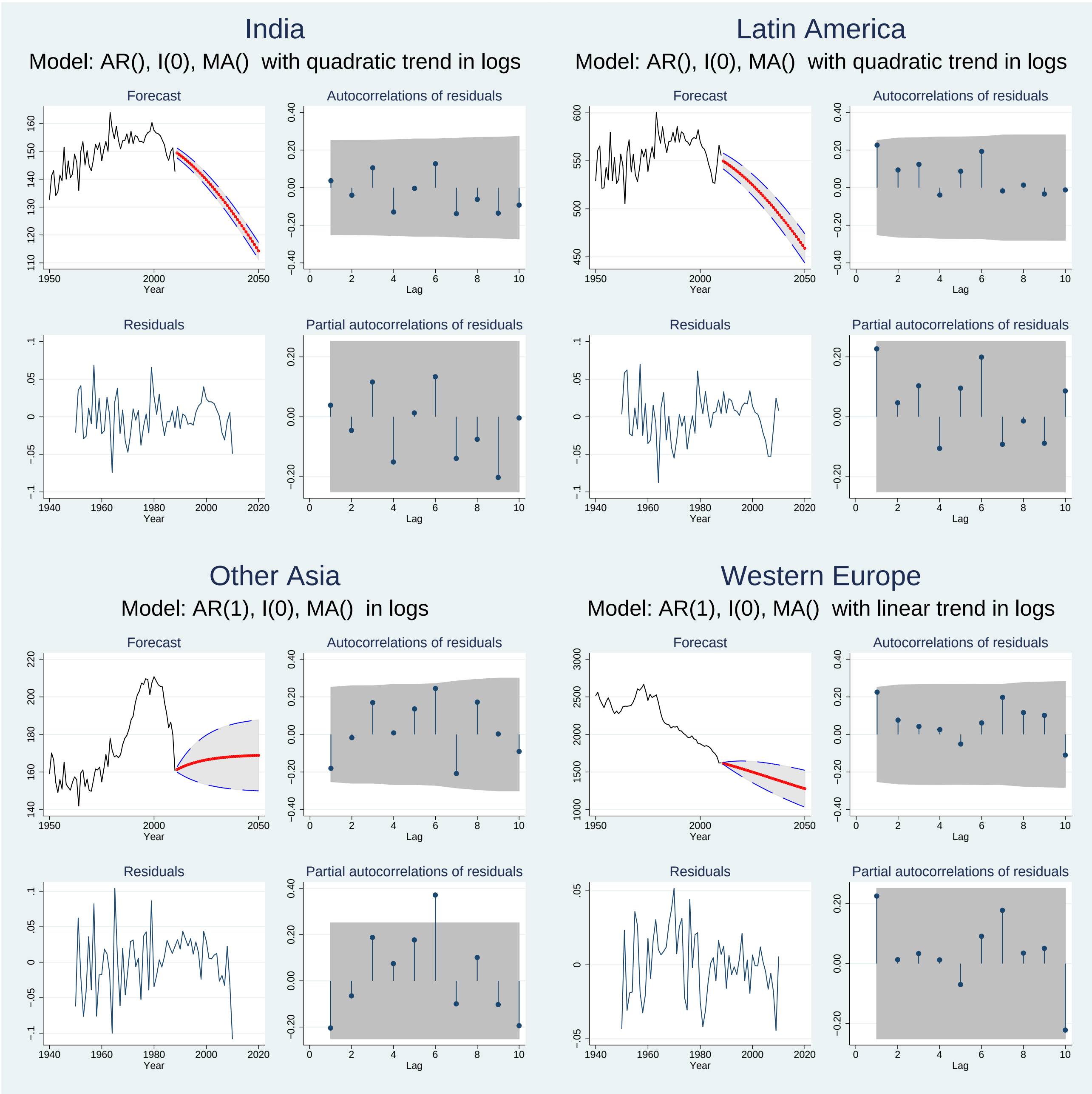


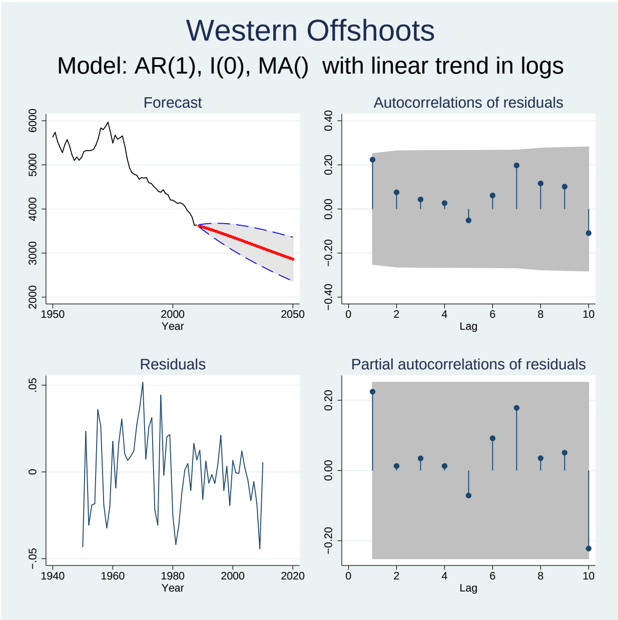

India
Model: AR(1), I(1), MA(1) in levels
Latin America
Model: AR(), I(1), MA() in levels
Other Asia
Model: AR(1), I(1), MA() in levels
Western Europe
Model: AR(1), I(1), MA() in levels
Forecast
Autocorrelations of residuals
Residuals
Partial autocorrelations of residuals
Year
Lag


Western Offshoots
Model: AR(1), I(0), MA() with quadratic trend in logs
Forecast
Year
Autocorrelations of residuals
Lag
Residuals
Year
Partial autocorrelations of residuals
Lag

Time Effects




India
Model: AR(), I(0), MA() with quadratic trend in logs
Forecast
Year
Autocorrelations of residuals
Lag
Residuals
Year
Partial autocorrelations of residuals
Lag
Latin America
Model: AR(), I(0), MA() with quadratic trend in logs
Forecast
Year
Autocorrelations of residuals
Lag
Residuals
Year
Partial autocorrelations of residuals
Lag
Other Asia
Model: AR(1), I(0), MA() in logs
Forecast
Year
Autocorrelations of residuals
Lag
Residuals
Year
Partial autocorrelations of residuals
Lag
Western Europe
Model: AR(1), I(0), MA() with linear trend in logs
Forecast
Year
Autocorrelations of residuals
Lag
Residuals
Year
Partial autocorrelations of residuals
Lag


Western Offshoots
Model: AR(1), I(0), MA() with linear trend in logs
Forecast
Year
Autocorrelations of residuals
Lag
Residuals
Year
Partial autocorrelations of residuals
Lag

Chapter 4

Corporate Governance, Environmental Regulations, and Technological Change

4.1. Introduction

The conventional view is that enhancing environmental quality via regulations hampers productivity growth by imposing extra costs on firms, such as reducing incentives to innovate. Mitigating this concern, there is a growing literature which provides empirical and theoretical evidence that more stringent environmental regulations direct R&D towards environment-friendly technologies (see surveys by Jaffe et al., 2003; Ricci, 2007; Popp et al., 2010). While ameliorating the negative effect on productivity, in an economy with profit maximizing firms, induced innovation may not totally offset the productivity loss, since it may crowd out some of the existing R&D activity (Popp and Newell, 2012). Indeed, if a given regulation could achieve to enhance overall productivity, it is expected that the profit maximizing firms would undertake it even in the absence of regulations. A limited number of empirical studies present opposing evidence that environmental regulations increase, not only R&D on clean technologies but "overall" R&D activity. For example, Jaffe and Palmer (1997) find that within industries there is a positive relationship between total R&D expenditure and stringency of environmental regulation proxied by pollution abatement costs. More recently, Hamamoto (2006) supports this finding by providing evidence from Japanese industries. In the current paper, we address these conflicting findings by departing from the profit maximizing manager assumption.

In the corporate governance literature, following Jensen and Meckling (1976), it is argued that, due to the agency problems associated with the separation of ownership and managership, firms may not be profit maximizer. As argued by Nelson and Winter

(1982) and Hart (1983), while maximization of profits is in line with the interests of the owners, firm behavior may also be determined by so called *satisficing managers* whose concerns are their private benefits. In Hart (1983), it is assumed that the managers get a fixed utility when their income satisfies a minimum level, and a utility of $-\infty$ otherwise. Under this condition, it is shown that owners may choose to pay a fixed wage as long as the firm achieves a threshold profit level, and zero otherwise. Another way to explain this situation is that managers obtain private benefits of control as long as they avoid bankcruptcy; however, they have a negative utility, if the firm is out of business (see Aghion and Griffith, 2005). We incorporate this approach in order to illustrate the implications of ownership-managership separation on the relationship between aggregate productivity growth and environmental regulation. We assume that a fraction of firms are governed by satisficing managers. The usefulness of this framework is that we are able to model corporate governance structure as a combination of two extreme cases, where owner controlled firms, maximizing their profits, are on the one extreme, and the mangerial firms, avoiding bankcruptcy, on the other. Therefore, we are able to present a general model leading to testable implications given the properties of existing ownership structure data.

Our model indicates that environmental regulation reduces the incentives of firms to innovate by eroding the monopoly rents for *owner-controlled firms*. However, for *managerial firms* where ownership and managership are separated, this conclusion is no longer valid. Since any effort beyond avoiding bankruptcy provides no extra utility for a satisficing manager, environmental regulations act as a disciplinary device, forcing them to innovate more. Depending on the prevalence of managerial firms in the economy, this distinction between owner-controlled and managerial firms makes it possible for environmental regulations to spur overall innovation, while ameliorating environmental pollution.

We conduct a country level empirical analysis in order to test our theoretical prediction that in countries where managerial firms with satisficing managers are more common, more stringent environmental regulations lead to higher innovation. Our main proxy for environmental stringency is tax-inclusive fuel prices across countries. As suggested by Aghion et al. (2010), fuel prices can be considered as the price of carbon, which may have an impact on innovative behavior in automobile firms. By using patent counts of countries as an indicator of innovative activity, we test whether fuel prices have a differential effect on innovation across countries depending on the relative prevalence of managerial firms. In order to construct a measure of ownership structure, we rely on firm level data. Since we use fuel prices as a proxy for environmental stringency, following Aghion et al. (2010), we restrict our sample to one industry category, "Manufacturing of motor vehicles, trailers, and semi-trailers" (hereafter, auto industry) as defined by the statistical classification in the European Union (NACE Rev 1.1 or NACE Rev 2). For this industry category, we classify firms as managerial or owner-controlled depending on their ownership concentration. Our panel data estimations show that in countries where the fraction of managerial firms is higher, increasing fuel prices have a higher innovation encouraging effect. This result suggests that by imposing more stringent environmental regulations, countries may achieve higher growth rates by encouraging innovation, as firms evolve to an ownership structure where separation of management from ownership is higher (Berle and Means, 1932).

This paper is related to the literature on the Porter Hypothesis. According to Porter and Van der Linde (1995), a properly designed environmental policy can lead to more innovation which reduces the net cost of regulation. They provide anecdotal evidence for how environmental regulations lead particular firms to innovate and adapt new technologies, which, in turn, results in a net benefit. In the literature, explanations of the Porter Hypothesis are centered around some type of market failure (see, among others, Simpson et al., 1996; Xepapadeas and de Zeeuw, 1999; Mohr, 2002; Greaker, 2003; Popp, 2005; Mohr and Saha, 2008; André et al., 2009). On the other hand, as in our paper, there are a small number of papers questioning the profit maximizing manager assumption (for example Gabel and Sinclair-Desgagné, 1997; Ambec and Barla, 2002). Among these, indicating a future research area, only the survey by Jaffe et al. (2003) discuss how the satisficing manager assumption can lead to outcomes in line with the Porter Hypothesis.

Secondly, our paper is related to the corporate governance literature investigating the effect of managership-ownership separation on the performance of firms, mostly measured by accounting profit rates or Tobin's Q (for example Demsetz and Lehn, 1985; Cho, 1998; Demsetz and Villalonga, 2001; Himmelberg et al., 1999; Holderness et al., 1998). Our paper differs from this literature by taking outcome of innovation, patents, as a measure of performance. The only paper taking this approach is **?**, which shows that there is a positive effect of institutional ownership on innovation. We also provide supportive evidence towards this hypothesis that fewer agency problems associated with ownership structure leads to more innovation. However, none of these papers investigate the differential effect of a given regulation depending on the ownership structure, which constitutes our main hypothesis in this paper. Our paper presents the only country level study on this relationship, which enables us to exploit directly the cross-country variations in environmental stringency.

There are only a limited number of papers investigating the relationship between patent counts and environmental regulations in a cross-country framework, such as Vries and Withagen (2005), Popp (2006), and Johnstone et al. (2010). All these papers focus on patenting activity in clean technologies in response to a specific type of environmental regulation. This is a weak form of the Porter Hypothesis, and there is also supportive empirical evidences at the firm level (for example Lanjouw and Mody, 1996; Brunnermeier and Cohen, 2003). On the other hand, our focus is the strong form of the Porter Hypothesis, predicting an increase in overall innovation in response to more stringent environmental regulations. Our goal is to identify the specific channel through the ownership structure, which may lead to outcomes in line with this hypothesis.

The rest of the chapter is organized as follows: In Section 2, we present our model. Section 3 describes our empirical strategy. We present the empirical results in Section 4. Section 5 concludes.

4.2. Model

In this section, in order to illustrate how environmental regulations can affect technological change depending on ownership structure, we present a static model of aggregate innovation. We choose the simplest framework in terms of the agent taxed for pollution (for example households or producers) and the stage of economic activity where pollution occurs (for example production or consumption), while maintaining that the model illustrates all our proposed channels through which environmental regulations affect innovation. However, our main theoretical results hold under different configurations.

4.2.1. Environmental Regulation and Innovation

There is a unique final output, Y, with price normalized to one. It is produced from a continuum of intermediate inputs indexed by $i \in [0,1]$, according to the aggregate production function:

$$Y = \int_0^1 A_i^{1-\alpha} x_i^{\alpha} di, \tag{4.1}$$

where x_i is the intermediate input used, and A_i is its productivity in final good production. The final good is used in consumption, intermediate goods production, and R&D activity. Therefore, our economy-wide resource constraint is given by

$$Y = C + X + Z,$$

where C is consumption, X is aggregate spending on intermediate goods production, and Z is total R&D spending. Assuming perfect competition in the final good market, each variety of intermediate inputs has to be paid equal to its marginal productivity,

$$p_i = \alpha A_i^{1-\alpha} x_i^{\alpha-1}, \tag{4.2}$$

where p_i is the price of variety i.

Market structure is assumed to be same for each intermediate good. The leading edge version of each variety is produced by a monopolist, and its production requires $1/\psi^M$ units of final good and $1/z$ units of pollution, which are constant and equal across i.[26] Therefore, the unit cost of each monopolist is

$$c^M = \psi^M + \tau z,$$

where τ is the economy-wide pollution tax levied per unit of pollution.[27] We assume that the monopolist is forced to charge a limit price in response to a competitive fringe of

[26] By taking emission intensity as a constant we exclude the possibility of induced innovation. We can also endogenize pollution intensity; however, we will see that our framework is sufficient to reach our main empirical hypothesis. On the other hand, endogenizing environmental innovations may lead to new and richer insights.

[27] Our model can be modified to allow for unit costs increasing with technology level, A_i. Thus, it can be extended to a case where unit cost changes across sectors.

imitators, producing the same variety with a higher marginal cost $\psi^F > \psi^M$.[28] Therefore, the monopolist sets the price of its variety equal to $c^F = \psi^F + \tau z$, which prevents a possible entry by the fringe firms. Substituting the limit price in equation (4.2) gives the demand for variety i:

$$x_i = \left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{1}{1-\alpha}} A_i. \tag{4.3}$$

Therefore, the monopoly profit in each sector is given by

$$\pi_i = (\psi^F - \psi^M)\left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{1}{1-\alpha}} A_i.$$

Any intermediate good producer can freely engage in R&D activity. Each R&D activity in sector i raises the productivity of variety (A_i) by a factor $\gamma > 1$, and grants a monopoly power to the innovator to produce the leading edge version of that variety. Following Aghion and Griffith (2005), we assume that R&D spending raises the size of innovation in a deterministic manner.

Producing a blueprint of a variety, which raises its productivity by size γ, requires to incur a variable cost of $\gamma^2 A_i/2$. Innovating firms differ in their degree of separation between ownership and managership. While owners benefit more from higher profits, the managers' main concern is their private benefit. We consider two extreme cases. On the one hand we have no separation, where the firm is a pure profit maximizer. On the other hand we have full separation, where there is a "conservative" firm in the sense that for the satisficing manager any effort beyond avoiding bankruptcy provides no further utility. We assume that innovation requires also to incur a fixed cost of $\kappa^j A_i$, where $\kappa^j > 0$. Here, $j \in \{pm, sm\}$ indicates the ownership structure, where pm stands for profit maximizer, and sm is for satisficing manager. The fixed cost of managerial firms is assumed to be higher compared to the profit maximizer firms, since managerial firms might have to incur extra costs in case of innovation such as training the manager. That is, $\kappa^{sm} \geq \kappa^{pm}$.

A profit maximizing manager chooses γ, which maximizes net benefit from R&D:

[28] Our main result is independent of the limit pricing assumption. However, it may be more intuitive in the context of the satisficing manager assumption. All results remain valid, if would assume a very high entry cost and the monopolist is free to charge the monopoly price

$$\max_{\gamma} \pi_{it}(\gamma A_i) - \frac{1}{2}\gamma^2 A_i,$$

The first order condition gives the optimal size of the innovation for the profit maximizing manager as

$$\gamma_i^{pm} = (\psi^F - \psi^M)\left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{1}{1-\alpha}}. \tag{4.4}$$

The monopoly profits can be rewritten in terms of γ_i^{PM} as $(\gamma_i^{PM})^2 A_i$. In order to have that an incumbent gains a non-negative benefit from innovation, $(\gamma_i^{pm})^2 A_i - (\gamma_i^{pm})^2 A_i/2 - \kappa^{pm} A_i \geq 0$ must hold. Rearranging this expression, we impose the following assumption:

Assumption 1: $\gamma_i^{pm} \geq \sqrt{2\kappa^{pm}}$.

Otherwise, there would be no incentive for the intermediate good producers to innovate.

The effect of pollution tax on a profit maximizing manager's innovation decision is given by following proposition.

Proposition 4.1. *More stringent environmental regulation discourages innovation by the profit maximizing firms. The rate at which the innovation size reduces is given by*

$$\frac{\partial \gamma_i^{pm}}{\partial \tau} = -\frac{z(\psi^F - \psi^M)}{\alpha(1-\alpha)}\left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{2-\alpha}{1-\alpha}} < 0. \tag{4.5}$$

Here, pollution tax increases the limit price, reducing the rents which the monopolist appropriates. In response, the monopolist reduces the innovation size, which, in turn, reduces the demand for intermediate goods. As a result, the monopolist's output decreases as well as the emission level. So, an improvement in the environmental condition comes at the cost of a lower output.

However, the innovation reducing effect of environmental regulation is not granted, once we take satisficing behavior of managers into consideration. A satisficing manager's objective is to avoid bankruptcy, and further effort reduces his utility. Therefore, the optimum R&D decision leads to zero profit net of the fixed costs:

$$\pi_i(\gamma A_i) = \kappa^{sm} A_i.$$

This gives the optimal size of innovation for the satisficing manager as

$$\gamma_i^{sm} = \frac{\kappa^{sm}(\psi^F + \tau z)^{\frac{1}{1-\alpha}}}{(\psi^F - \psi^M)\alpha^{\frac{1}{1-\alpha}}}. \tag{4.6}$$

The following proposition states the effect of pollution tax on a satisficing manager's innovation decision.

Proposition 4.2. *More stringent environmental regulation encourages innovation by satisficing managers. The rate at which the innovation size increases is given by*

$$\frac{\partial \gamma_{it}^{SM}}{\partial \tau} = \frac{z\kappa(\psi^F + \tau z)^{\frac{\alpha}{1-\alpha}}}{(\psi^F - \psi^M)(1-\alpha)\alpha^{\frac{1}{1-\alpha}}} > 0. \tag{4.7}$$

As in the profit-maximizing case, pollution tax reduces the monopoly rents, which, in turn, reduces the demand for intermediate goods. If the satisficing manager responds by reducing innovation, bankruptcy is unavoidable, since a satisficing manager always chooses innovation size at a level just preventing bankruptcy. Therefore, a satisficing manager increases innovation size in order to maintain the same level of demand before the increase in pollution tax. That is, environmental regulations reduce slack behavior in satisficing managers in the sense that they are forced to have higher innovation size in order to avoid bankruptcy.

A peripheral point in this model, which does not have any consequences on the main results is that the relative innovation size of two types of firms depends on the difference between their fixed costs. When $\kappa^{sm} = \kappa^{pm}$, profit maximizer has a higher innovation size. When κ^{sm} is higher, the difference in the innovation size of two types of firms reduces. When κ^{sm} is sufficiently high, the innovation size of a satisficing manager is suboptimally higher than that of a profit maximizer. Such a case may seem unrealistic in our model, where κ reflects the fixed costs of innovation to the firms. However, avoiding bankruptcy is an extreme type of satisficing behaviour. In a more detailed model, one can model satisficing behaviour by specifying the utility function of a manager, and excluding fixed costs of R&D. In this case, κ can be related to the manager's ability of adopting to a new technology. Or it might be related to some career concerns. Therefore, the innovation size of a satisficng manager can be separated from the cost structure of the firm given that satisficing behaviour always avoids bankruptcy. In such a model, a suboptimally high innovation by a satisficing manager is not unrealistic. Note that the relative innovation size of two types of firms does not effect our main results which are regarded with the response of firms to a pollution tax. Therefore, in the current model, we avoid such complications.

4.2.2. Aggregation

The pollution level resulting from production of a variety is given by $E_t = zx_i$. Substituting the equilibrium value of x_i gives the equilibrium pollution by sector i:

$$E_i = z\left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{1}{1-\alpha}} A_i.$$

Obviously, there are two channels through which raising τ affects E_i. The direct effect is a decline in monopoly rents which reduces equilibrium pollution. The indirect effect is through its influence on productivity. For the profit maximizer, this indirect effect is also negative. Therefore, pollution decreases when environmental regulation is more stringent. However, for the satisficing manager, the indirect effect is positive. It can be shown that these two opposing effects are neutralized. Hence, satisficing managers keep the equilibrium amount of intermediate good unchanged. Therefore, the pollution level does not change. In order to clarify this point, it is convenient to work with the pollution-output ratio. Substituting equation (4.3) in (4.1), the final good produced by using variety i is equal to

$$Y_i = \left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{\alpha}{1-\alpha}} A_i. \tag{4.8}$$

Note that the effect of τ on Y_i is similar to its effect on E_i. The emission-output ratio can be calculated as $z\alpha/(\psi^F + \tau z)$. Now it is clear that the emission-output ratio is independent of the degree of separation between ownership and managership, and more stringent environmental regulation always reduces this ratio. The distinction is that in response to a higher pollution tax, in order to reduce this ratio, profit maximizing firms decrease both output and emission, while satisficing managers increase output to meet this ratio. In other words, the satisficing manager's optimal response is only to increase productivity of their variety, so that the negative direct effect of taxation through depressing monopoly rents is balanced with an increase in demand. The result is an unchanged level of pollution, but a higher output level through productivity gains.

The equilibrium final good production can be calculated by integrating both sides of equation 4.8 which leads to

$$Y = \left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{\alpha}{1-\alpha}} \bar{A} = \left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{\alpha}{1-\alpha}} \left(\underbrace{\int_0^s A_i di}_{A^{sm}} + \underbrace{\int_s^1 A_i di}_{A^{pm}} \right).$$

where the fraction of firms with satisficing managers is denoted by $s \in [0,1]$. The average technology levels are denoted by $\bar{A}$ for the whole economy, A^{sm} for the type sm firms, and A^{pm} for the type pm firms. Therefore the average innovation size, γ^A, is given by $\gamma^A \bar{A} = (1-s)\gamma^{pm} A^{pm} + s\gamma^{sm} A^{sm}$.[29] Substituting γ^{pm} and γ^{sm} yields

$$\gamma^A = (\psi^F - \psi^M) a^{pm} \left(\frac{\alpha}{\psi^F + \tau z}\right)^{\frac{1}{1-\alpha}} + s \left[\frac{\kappa^{sm} a^{sm} (\psi^F + \tau z)^{\frac{2}{1-\alpha}} - (\psi^F - \psi^M)^2 a^{pm} \alpha^{\frac{2}{1-\alpha}}}{(\psi^F - \psi^M) \alpha^{\frac{1}{1-\alpha}} (\psi^F + \tau z)} \right], \tag{4.9}$$

where $a^j = A^j/\bar{A}$. Here, the first term represents the direct effect of τ on innovation size, γ^A, which is negative. This direct channel is about decreasing monopoly rents, and independent of ownership structure. The second term represents the indirect effect through the ownership structure, s, which is positive. The total effect of τ on γ^A is equal to $(1-s)(\partial\gamma^{pm}/\partial\tau)a^{pm} + s(\partial\gamma^{sm}/\partial\tau)a^{sm}$, which has an ambiguous sign. We can directly substitute the partial derivatives in this expression from equation (4.5) and (4.7) which lead to following expression:

$$\frac{\partial \gamma^A}{\partial \tau} = \tilde{\alpha} z \left(\psi^F + \tau z\right)^{\frac{\alpha}{1-\alpha}} \left[-(1-s) \left(\psi^F + \tau z\right)^{-\frac{2}{1-\alpha}} a^{pm} + s \left(\frac{\kappa^{sm} a^{sm}}{((\psi^F - \psi^M)\tilde{\alpha})^2} \right) \right], \tag{4.10}$$

where $\tilde{\alpha} = \alpha^{1/(1-\alpha)}/(1-\alpha)$. Taking the derivative of equation (4.10) with respect to s leads to our main result:

Proposition 4.3. *The derivative of equation (4.10) with respect to s is given by*

$$\frac{\partial^2 \gamma^A}{\partial \tau \partial s} = \tilde{\alpha} z \left(\psi^F + \tau z\right)^{\frac{\alpha}{1-\alpha}} \left[\left(\psi^F + \tau z\right)^{-\frac{2}{1-\alpha}} a^{pm} + \frac{\kappa^{sm} a^{sm}}{((\psi^F - \psi^M)\tilde{\alpha})^2} \right] > 0, \tag{4.11}$$

[29] We measure the degree of separation in an economy as a linear combination of full separation and no separation. Here, the satisficing manager assumption is very useful, in the sense that it illustrates an extreme case where the agency costs are most severe, in contrast to the profit maximizing firm assumption. As indicated by Aghion et al. (1997), in practice, the objective function of each individual manager is likely to be a convex combination of these two extreme cases. Therefore, our assumption, which allows the construction of an indicator for a country level study, can be seen as an approximation.

which is always positive. Therefore, more stringent environmental regulations marginally lead to relatively higher innovation in economies where the degree of separation between ownership and managership is relatively high.

These results suggest that, everything else constant, for the countries where managerial firms are more prevalant, higher pollution taxes have a higher innovation encouraging effect, although their performance in terms of innovative outcome can be lower. The total effect of an increase in pollution tax depends on the magnitude of two countervailing effects. The direct effect is negative; however, this is mitigated by its effect through ownership structure.

According to equation (4.11), compensation of the direct negative effect of fuel prices is stronger when emission intensity is higher.[30] If the indirect positive effect of fuel prices in relation to ownership structure is sufficiently strong, a positive net effect of pollution tax on innovation is also possible. Such an outcome is more likely when fraction of managerial firms, fixed costs of R&D to satisficing managers, competitive pressure from the fringe firms, emission intensity, and relative knowledge stock of satisficing firms compared to owner-controlled firms are all higher.

It is important for our empirical strategy that this result of our model is independent of the type of agent taxed for environmental pollution. If the tax is levied on households or final good producers instead of intermediate good producers, demand for intermediate goods falls. The monopolists' response depends on the type of ownership. That is, the indirect channel works in the same manner with the case where the monopolists are taxed. Therefore, in all cases, there is a differential effect of pollution tax depending on the ownership structure. However, taxing households or final good producers does not trigger the direct channel which works through reducing monopoly rents.

Our goal in this section is to illustrate the proposed mechanism, and derive consistent testable predictions. Therefore, our model is kept simple, at the cost of leaving some interesting questions unanswered. Importantly, we do not analyze under which conditions, regulations might not create this differential effect on innovation. More specifically, in our model, any regulation that triggers a change in the demand for the polluting good will result in the same differential innovative response depending on ownership structure.

[30] A sufficient condition for this result is $(2-\alpha)/(1-\alpha) > \tau + \psi^F/z$. That is, emission intensity should not be too small.

Indeed, this is not realistic, and the question of why some regulations may create this outcome, while some may not, is an interesting question for future research. One possible avenue relates to the assumptions on the substitution possibilities. When there exists a substitute for the input which is taxed, both the satisficing managers and the profit maximizing owners may switch between the substitutes. The net effect on innovation will depend on the structure imposed on the variable and fixed costs. Therefore, the mechanism proposed may not work in every circumstance. Our underlying assumption is that either the elasticity of substitution is low, or the tax rates are not sufficiently high, preventing a switch. These assumptions make sense in the current context. First of all, although there have been substantial improvements in clean technologies such as non-renewable energy production, it is still in its infancy compared to dirty technologies, leading to a low substitution possibility. Secondly, there are well-known political frictions (for example international coordination failures) such that environmental regulations cannot be as stringent as required in order to trigger a switch towards clean technologies. However, these assumptions may not be argued to be valid in every context.

Another possible improvement is to model the behavioral consequences of managership-ownership separation more precisely. Satisficing and profit maximizing managers assumptions are very useful in the same framework, in the sense that they illustrate extreme cases in terms of agency costs. However, one can go further by endogenizing ownership structure. This may have interesting implications, since in response to a given regulation, firms may re-evaluate their ownership structure.

4.3. Empirical Strategy

The main predictions of our theoretical model are

1. Direct effect of environmental stringency on innovation is negative.

2. Indirect effect of environmental stringency through ownership structure stimulates greater innovation when ownership and managership are more separated.

In order to test these predictions, we follow the common approach of using patent counts as a proxy for innovation size. For count variables, the outcome is non-negative, mostly zero for a substantial number of observations, and unbounded from above. This requires us to use appropriate estimation techniques accounting for these properties. We will show that our patent count data is also characterized by these properties. Therefore, we estimate the hypothesized relationship using non-linear count data estimation techniques in a maximum likelihood framework. Consider the following non-linear specification:

$$E(INN_{it}|Z_{i,t-1}) = \exp(\alpha_1 STR_{i,t-1} + +\beta OWN_{i,t-1} STR_{i,t-1} + \delta X_{i,t-1}), \tag{4.12}$$

where INN is the innovation size of country i in year t, and $Z_{i,t-1}$ stands for all right hand side variables. In equation (4.12), STR is a measure of environmental stringency; therefore, our first prediction indicates that α_1 should be negative. OWN is the fraction of firms with satisficing managers in country i. Therefore, β measures the indirect effect of STR through OWN. Our second prediction indicates that it should be positive. Finally, X is a set of control variables including ownership indicator, country-fixed effects, and a full set of time dummies.

For environmental stringency, the ideal measure is the shadow price of pollution, which is difficult to observe. The common approach is to use proxies such as polluting-input prices, abatement expenditures, or characteristics of environmental regulations (Jaffe et al., 2003). Since the policy variable in our model, measuring environmental stringency, is a pollution tax, using polluting-input prices is preferable as a proxy for environmental stringency. Following Aghion et al. (2010), we use tax-inclusive fuel prices as our main proxy for environmental stringency.[31] Fuel prices can be considered as a proxy for the shadow price of pollution (Aghion et al., 2010), which refers to the pollution tax, τ, in our theoretical model. Of course, not all industries are sensitive to tax-inclusive fuel prices in terms of innovative behavior. However, it is expected that, when fuel prices increase, automobile manufacturers will respond by making carbon saving innovations

[31] In our theoretical model, tax is levied on the innovating sector; however, note that fuel prices and taxes are end-use prices and taxes, and paid by households or enterprises. Hence, these cannot be directly considered as polluting-input prices for the intermediate sector. However, as we discussed previously, our main theoretical predictions remain robust, if τ is levied on households or final good producers instead of intermediate good producers. Furthermore, we make no assumptions about the stage of economic activity at which pollution occurs. Our only assumption is that pollution is directly linked to the equilibrium amount of intermediate goods.

(Aghion et al., 2010). Therefore, we restrict our sample to the auto industry. We also check the robustness of our results by using R&D subsidies proxied by government R&D budget appropriation and expenditure on environmental protection.

In order to construct the ownership indicator, implied by our theoretical model as a fraction of ownership-managership separated firms, we rely on firm level data. We follow the general approach of using ownership concentration as a proxy for ownership-managership separation, since a higher level of concentration reduces agency problems. Adopting this approach, one strand of literature investigates the determinants of ownership structure and its effect on firm performance, following Demsetz and Lehn (1985), where ownership structure is proxied by the share of dominant shareholders such as the top five or top twenty shareholders. However, such a measure can be criticized for not taking into account the fact that managers of many firms hold some shares, which mitigates agency problems. For this reason, a number of studies prefer to use managerial shareholdings instead of ownership concentration (Cho, 1998; Hermalin and Weisbach, 1988; Himmelberg et al., 1999; Holderness et al., 1998; Loderer and Martin, 1997; McConnell and Servaes, 1990; Morck et al., 1988). However, Demsetz and Villalonga (2001) argue that this measure also has limitations. The most important problem is that dominant shareholders generally have representatives as board members, and they cannot be considered as satisficing managers. Revealing the severity of such a problem, Demsetz and Villalonga (2001) report that in their sample there is a correlation of 0.67 between the fraction of shares held by families and the fraction of managers' shares. Clearly, both proxies are imperfect; however, using dominant shareholder data has some additional advantages in our case, which we will discuss later.

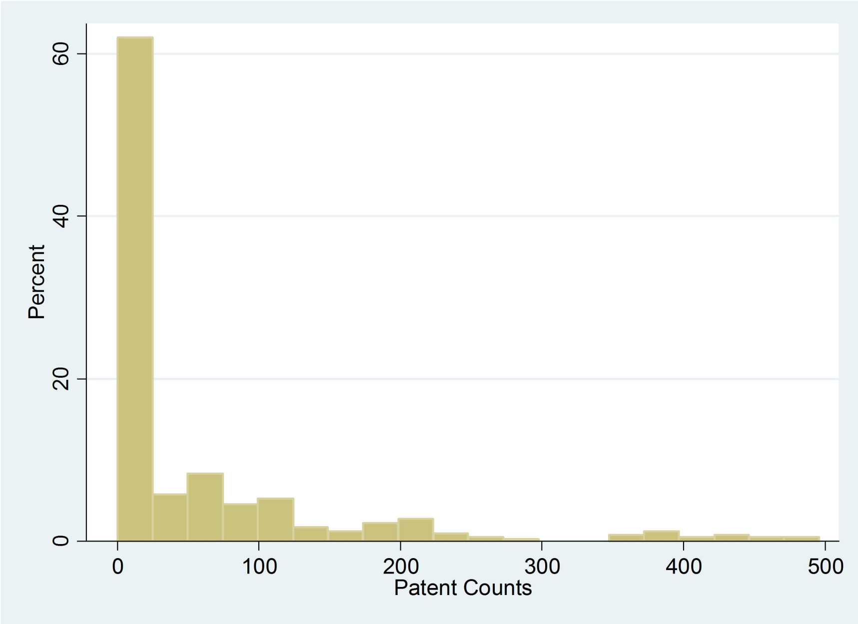
4.3.1. Data and Descriptive Statistics

Patent Data

The World Patent Statistical Database (PATSTAT) by the European Patent Office (EPO) provides data on firm level patent counts which are categorized according to the international patent classification (IPC). The statistical office of the European Union (EUROSTAT) documents the applications to the EPO at national level by priority year and NACE industry classification.[32] It covers an unbalanced panel of patent counts for

[32] NACE classes are derived in line with IPC categories.

Figure 4.1: Distribution of Patent Counts



Note: Graph excludes the range where patent counts are higher than 600 for which there are very few observations.

48 countries from 1977 to 2009. We extract all available Eurostat data for the automobile industry.

The number of patents is a count variable, including only non-negative values with no natural upper bound. A common property of count variables is that they are highly skewed to the right. As can be seen in Figure 4.1, these properties also hold for our patent count variable. About 60% of the observational values are close to zero, more specifically below 20. This is crucial for our estimation strategy, which we will discuss in detail in following sections.

Count of patents is a preferred measure of innovative outcome and activity of a firm or country for several reasons, which are all well documented in the literature (see **?** for a detailed discussion). However, it is not a perfect measure, and there are some important points that should be taken into account in an econometric analysis employing patent counts. First of all, the propensity of an inventor to apply for a patent to EPO may vary across countries depending on the market conditions, quality of institutions in protecting the patent rights, and variation in patent law. In order to deal with these country specific differences, we simply employ country fixed effects in our regressions.

Another problem is that patent counts weight all patents equally, ignoring their relative importance. The common practice in the literature, dealing with this problem, is to make robustness checks by using a measure of patent quality in order to obtain quality adjusted patent counts. For this purpose, future citations to a patent is a frequently used measure of its quality. Obviously, this requires a dataset with a sufficiently long time dimension in order to allow sufficient time for the patents observed in the final period to receive citations. For this reason, such an approach is not feasible in our case, since our ownership data restricts our sample between 2003 and 2009. Still, there are several reasons why unweighted patent counts can be sufficient in our case. Firstly, this problem is partially mitigated in our study by only using patent applications to the EPO. Since an application to the EPO is more expensive compared to an application to national patent offices, this strategy eliminates low quality patents (Johnstone et al., 2010). Secondly, in a cross-country setting, using citations in order to weight patents may not be so crucial compared to a firm level study, since the change in mean quality at country level is not expected to be as high as the change within the firms. The reason is that a firm can change its innovation quality easily, for example by hiring more qualified researchers. However, at the country level, such an attempt requires long term investments, such as education. Therefore, using patent counts may suffice, considering our short time dimension of our dataset.

Our estimation strategy can still account for some source of variation in the quality distribution of patents. Firstly, if the quality differences among countries are due only to country specific time-invariant characteristics, then fixed effects estimation can control for this concern. Moreover, if the change in quality is a worldwide trend, then time dummies eliminate this effect. Therefore, the assumption required is that, in our time period, relative quality differences among countries does not change over time. More specifically, we assume that, between 2003 and 2009, the relative differences in the citations received by the representative patents of countries remain constant. This can be argued to be valid due to our short time period.

This assumption is also consistent with recent theoretical and emprical evidence. In theory, there are two reasons to expect a systematic variation in the quality of innovative activity. Firstly, a variation in quality may be observed over time, such that as knowledge stock increases, it may be more difficult to create new inventions. Note that, in theory, this suggests that we have to control for initial levels of technology. Fixed effects estimation controls for such a concern. However, in practice, such diminishing returns may exhibit heterogenous time patterns across countries. However, Popp (2012) finds no emprical evidence for this. Secondly, it may be argued that the highest quality projects are those first undertaken by entrepreneurs, and additional investment goes to lower quality projects. Therefore, one might expect to see diminishing returns to R&D spending in a given year. Indeed, Popp (2012) finds that an increase in the number of patents in a given year decreases the probability that these patents will receive future citations. As a consequence, higher patent counts are associated with lower quality. Therefore, ignoring citation weights requires us to assume that the rate of diminishing returns in terms of innovation quality does not exhibit heterogenous time patterns across countries, which seems realistic due to our short time period. In this case, again, such a concern can be accounted for by fixed effects estimation with time dummies. Note that it is realistic even to assume that this rate does not change over short time periods, since, as we argued, a shift in the quality distribution of innovative outcome at country level requires long-term investments such as education. This may also explain the fact that most results in the literature do not change by weighting patent counts with citations.

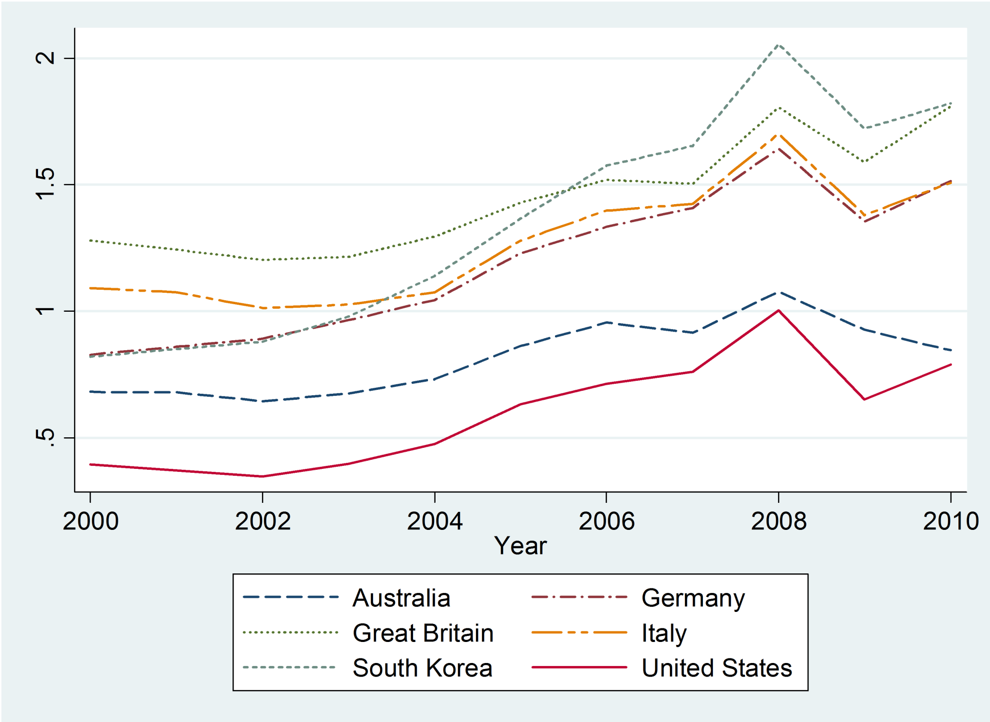
Tax Inclusive Fuel Prices

Fuel price data are from the International Energy Agency (IEA) which are available for 35 countries from 1978 to 2011. Our preferred variable is tax-inclusive diesel-fuel prices, since it reflects the full price of using carbon.[33] The IEA provides this data for different sectors and types of fuels. Among these, we use household data, since it is the largest. We present robustness checks with the alternative measures.

Figure 4.2 presents the tax inclusive fuel prices for some auto producer countries from distinct locations around the world. There seems to be a worldwide general upward trend with a peak around 2008. However, there is also heterogeneity in the patterns exhibited, in which the different tax regimes of these countries may play a role.

[33] Market may respond to tax exclusive fuel prices and taxes differently. However, using tax data reduces our sample size substantially.

Figure 4.2: Tax Inclusive Fuel Prices (US$/unit, PPP adjusted)



Ownership Data

As a measure of separation between ownership and managership, we use the Bureau van Dijk's (BvD) independence indicator which is available in the ORBIS database. BvD's indicator classifies firms into four categories: A company is classified as A when no recorded shareholder has more than 25% of the shares. If a company is not in category A, and there is no recorded shareholder with an ownership percentage over 50%, then it is classified as B. Companies in categories C and D have at least one owner with an ownership percentage over 50%. The final category is U, indicating unknown situation. We consider category A and B firms as managerial, and category C and D firms as owner-controlled. We also control for sensitivity of our results by taking only category A firms as managerial.

We filter the data of automobile firms from 48 countries for which patent counts are available in the EUROSTAT database. This produces a list of 86,015 firms, which is close to the total population size. As can be seen in Figure 4.3, among these firms, we have around 13,000 active firms per year with a known ownership situation for 2005 and onwards, and around 8,000 from 2003 to 2005. Since the number of firms with a known ownership situation remains very small until 2003, we exclude the years before this date.

Figure 4.3: Total Number of Firms by Ownership Category

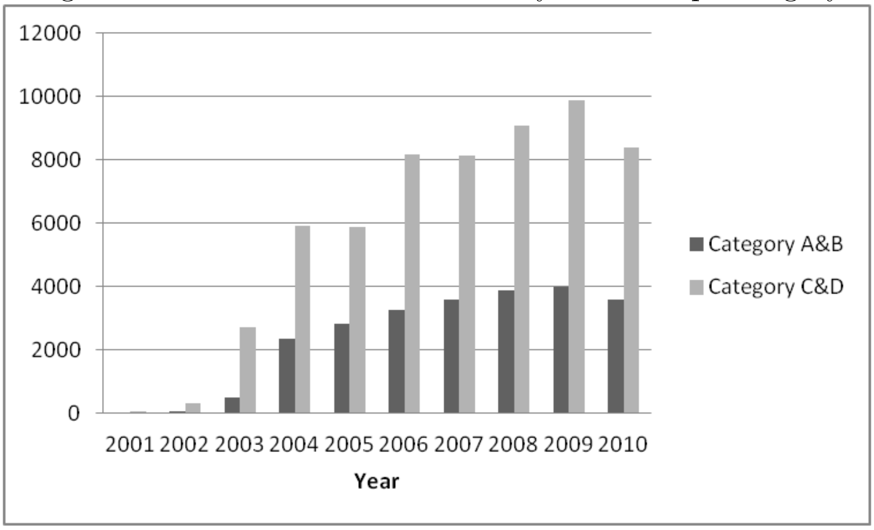
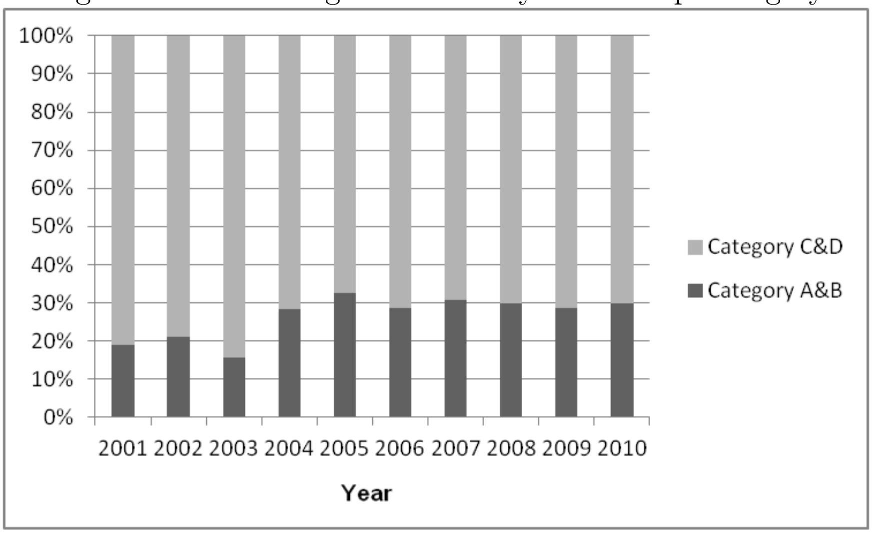


Figure 4.4: Percentages of Firms by Ownership Category



Although there are a substantial number of observations, this number increases significantly over time, which may constitute a problem if the selection of observational units for an individual country changes over time depending on firm characteristics affecting the composition of the ownership structure. Figure 4.4 illustrates that the relative number of firms across categories (for all countries in our sample) changes little over time. Interestingly, this pattern is also valid for 2001 and 2002 where the number of firms is relatively small compared to the following years. Here, it seems unlikely that the problem we described and the pattern depicted in Figure 4.4 coexist. In Section 4, we extensively investigate whether sample selection raise a problem in our analysis.

The firms with no ownership information at a specific year are denoted with "U" by the BvD indicator. Although the BvD dataset lists nearly all firms for most of the countries in our sample, for some countries, the firms in category U constitute a high fraction of the listed firms. As long as the firms with a known situation do not differ across countries in those characteristics, which are correlated with ownership structure, this will not constitute a problem. However, it is possible that in some countries only the large firms have a known situation, while in some other countries, the unknown situation is more heterogeneous in terms of firm size. At the same time, firm size may be related with ownership structure as argued in the corporate governance literature, following Demsetz and Lehn (1985). In order to control for such a situation, we employ the size data, which is available in the ORBIS database for all the listed firms, including those with unknown ownership status. BvD categorizes firms as "very large", "large", "medium", and "small". This categorization depends on pre-determined thresholds for four variables: operating revenue, number of employees, total assets, and being publicly listed or not. For example, a firm is categorized as very large, even if only one of these variables exceeds the relevant threshold. Therefore, prevalence of one of these variables is sufficient to determine the size of a firm, which explains why it is available for all listed firms.

There are several advantages of using the BvD independence indicator. Firstly, it can be constructed for a substantially larger number of firms compared to the measures that depend on the top five or twenty shareholder ownership percentage. Basically, there are two reasons for this. Firstly, what we need is not the shares of the top five shareholders, but only information about owners who hold 50% of the shares. That is, when two shareholders have 50% ownership, then other unknown shareholders will not have a share higher than 50%. Secondly, for many firms, not all shareholders are recorded, and those recorded are more likely to be dominant ones. That is, it is very likely that other unknown owners' shares are less than the shares of the recorded ones. Indeed, BvD indicates the reliability of the indicator with a "+" or "−" sign. For example, if a firm is classified as A−, this means that there is only one recorded shareholder whose ownership percentage is less than 25%. As a result, our sample size is substantially larger than the previous studies, which mostly focus on publicly listed or very large companies. As Aghion and Griffith (2005) argue, when the focus is on the differential behavior between owner-controlled and managerial firms, restricting the sample with large or publicly listed firms may not be appropriate, since these firms are likely to be managerial. This may have a substantial effect in our case, since our goal is to construct a country level indicator which is the fraction of managerial firms. For example, although large firms from developed countries are more likely to be managerial firms compared to their counterparts in developing countries, it is also more likely that the fraction of small size firms which are likely to be owner-controlled is also higher in developed countries. Therefore, having a representative sample is crucial.

Another advantage of using BvD's indicator is that it accounts for an indirect ownership structure. Consider a case where firm A owns 80% of firm X and 80% of firm Y, and both firms X and Y have 40% of shares of firm B. Here, if we only consider direct ownership linkages of B, it is classified as a managerial firm, since neither X nor Y has more than 50% of shares. However, A indirectly owns 64% of B. So, it is classified as owner-controlled when we consider the indirect ownership structure. As Demsetz and Villalonga (2001) argue, such external owners mostly have representatives in firm boards, who cannot be considered as "pure manager personnel".

Other Data

In order to check the robustness of our results, we also use proxies for R&D subsidies on environmental issues which is expected to relax the environmental constraints on firms, as opposed to the tax-inclusive fuel prices. Our first measure of R&D subsidy is the percentage of government R&D expenditure in overall government expenditures, which is extracted from the EUROSTAT database. However, using this data shrinks our already small sample size (country level) substantially. For this reason, as our preferred measure, we employ the percentage of government R&D budget appropriations on environmental issues in overall government expenditure, which is also extracted from the EUROSTAT database. Using budget appropriations leads to a number of observations about two times higher than that obtained by using expenditure data.

4.3.2. Estimation Strategy

As we illustrated in Figure 4.1, our dependent variable, the number of patents, exhibits the typical count data properties, a highly right-skewed distribution with no natural upper bound, and with majority of observations close to zero. A common approach applied in such a situation is to perform a Poisson maximum likelihood (PML) estimation, which assumes that the dependent variable has a Poisson distribution. On the other hand, the drawback of the PML estimation is the assumption of equal conditional mean and variance, which is rejected in most cases, since count data are generally overdispersed. All our estimations will employ individual specific fixed effects, which mitigates the concern about overdispersion. In case this is not sufficient, we fully account for a potential overdispersion problem by using the Poisson quasi maximum likelihood (PQML) estimator with cluster robust standard errors, which is free of the conditional variance specification. This approach is fully efficient among the estimators which only use the conditional mean assumption (see **?**). As a result, our estimations are robust to a potential misspecification of the conditional variance.

Another method directly accounting for overdispersion is the negative binomial maximum likelihood (NBML) estimation, which assumes that conditional variance is a quadratic function of the conditional mean. If this specification is correct, then the NBML is more efficient than the PQML estimation. Therefore, we also employ the NBML estimation strategy.[34] These three estimation strategies account for most of the problems raised for count data applications, and are sufficient in most applications (**?**).

[34]In the Appendix, we also present robustness checks for the negative binomial estimations by calculating the standard errors using the Jack-knife method.

4.4. Empirical Results

4.4.1. Baseline Estimations

Our baseline Poisson estimations are presented in Table 4.1, where all regressions incorporate the logarithm of GDP per capita, a set of time dummies, and country fixed effects. We start with PML estimations. Column (1) indicates that the effect of lagged prices on patent counts is significant with a negative sign. The sign and significance of this coefficient is robust to include other variables of interest in the following columns. This finding confirms our first prediction, that an increase in the shadow price of pollution discourages innovation by reducing the monopoly rents of firms. This channel works independently of the ownership structure.

Table 4.1: Baseline Poisson Regressions

Dependent variable: Patent counts

	Poisson ML			Poisson QML		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Fuel price	-2.210***	-2.017***	-2.333***	-2.210***	-2.017***	-2.333***
$\ln(1 + P_{i,t-1})$	(0.295)	(0.328)	(0.339)	(0.493)	(0.523)	(0.375)
Ownership structure		0.079	-0.616***		0.079	-0.616***
$Ownstr_{i,t-1}$		(0.065)	(0.196)		(0.086)	(0.199)
Interaction			1.080***			1.080***
$\ln(1 + P_{i,t-1}) * Ownstr_{i,t-1}$			(0.288)			(0.382)
ln(GDP per capita)	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
Years	2003-09	2003-09	2003-09	2003-09	2003-09	2003-09
AIC	1117	1075	1062	1117	1075	1062
BIC	1140	1100	1091	1140	1100	1091
Ave. obs. per group	5.7	5.5	5.5	5.7	5.5	5.5
Num. of countries	34	34	34	34	34	34
Observations	194	186	186	194	186	186

Standard errors are in parentheses. Significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In column (2), we also include the lagged ownership variable. The estimated coefficient is insignificant. Finally, in column (3), we add the interaction term into the regression. Its effect on patent counts is highly significant. This confirms our third prediction, that imposing higher environmental costs on firms has a differential impact depending on the ownership structure. The sign of this coefficient is positive, which suggests that when managerial firms are more prevalent, environmental regulation encourages innovation more. In this full specification, the direct effect of ownership structure turns out to be highly significant with a negative sign. This supports the theoretical arguments in the corporate governance literature claiming that owner controlled firms may perform better in the absence of regulation.

The estimated coefficients in Table 4.1 are interpreted as the elasticities. In column (3), the estimated direct effect of fuel prices on innovation is -2.333. However, the size of this negative effect substantially reduces in absolute terms due to the indirect effect in relation to the ownership structure, which leads to a net effect of -1.253. We also calculate the marginal effect of fuel prices at $OWN = 1$ and for the maximum value of fuel price, leading to a net marginal effect of fuel price as low as -0.18. Therefore, according to our estimation, the compensating effect of the ownership channel is substantial. However, there is no evidence for the theoretical possibility that the fuel prices might increase innovation.

We replicate these regressions by using the PQML estimation strategy with cluster-robust standard errors. Results are presented in columns (4) to (6). In terms of sign, size, and significance of the coefficients, these results confirm the findings from the PML estimations. The increase in the standard errors is not large, which indicates that the overdispersion problem may not be very severe.

Alternatively, negative binomial estimation can directly account for overdispersion. Therefore, we replicate these regressions by using negative binomial estimations with fixed effects. The results, presented in Table 4.2, confirm the findings from the Poisson estimations in terms of sign, size, and significance of the coefficients. The Akaike and Bayesian information criteria indicate that negative binomial estimations have a slightly better fit than the Poisson estimations. This also indicates that the overdispersion problem is not likely to be severe.

Table 4.1 also reports the number of observations subject to the baseline regressions.

Table 4.2: Baseline Negative Binomial Regressions

Dependent variable: Patent counts			
Independent variables	(1)	(2)	(3)
Fuel price	-2.175***	-1.974***	-2.344***
$\ln(1 + P_{i,t-1})$	(0.406)	(0.453)	(0.429)
Ownership structure		0.076	-0.606**
$Ownstr_{i,t-1}$		(0.095)	(0.267)
Interaction			1.011***
$\ln(1 + P_{i,t-1}) * Ownstr_{i,t-1}$			(0.371)
log(GDP per capita)	yes	yes	yes
Country fixed effects	yes	yes	yes
Time dummies	yes	yes	yes
Years	2003-09	2003-09	2003-09
AIC	1097	1055	1050
BIC	1123	1084	1082
Ave. obs. per group	5.7	5.5	5.5
Num. of countries	34	34	34
Observations	194	186	186

Note: Standard errors are in parentheses. Significance is indicated as

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

It is noteworthy that, despite our small number of observations, we have highly significant results. Following, we perform various robustness checks on these results. Hereafter, we will not report the results from the PML estimations, by noting that all our results from the PQML estimations does not change with the PML, since the PQML only adjust the standard errors of the PML in the upward direction.

4.4.2. Robustness Checks

Setting a Minimum Number of Observations to Calculate Ownership Indicator

Table 4.3: Setting a Minimum Number of Observations

Dependent variable: Patent counts

	Poisson QML			Negative Binomial		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Fuel price	-2.385***	-2.373***	-2.378***	-2.456***	-2.493***	-2.500***
$\ln(1+P_{i,t-1})$	(0.391)	(0.411)	(0.405)	(0.431)	(0.434)	(0.436)
Ownership structure	-0.671***	-0.740***	-0.726***	-0.667**	-0.682**	-0.659**
$Ownstr_{i,t-1}$	(0.191)	(0.194)	(0.191)	(0.280)	(0.278)	(0.279)
Interaction	1.194***	1.422***	1.400***	1.152***	1.377***	1.343***
$\ln(1+P_{i,t-1}) * Ownstr_{i,t-1}$	(0.395)	(0.373)	(0.371)	(0.395)	(0.396)	(0.398)
ln(GDP per capita)	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
Firm num. per data point	>20	>40	>60	>20	>40	>60
Years	2003-09	2003-09	2003-09	2003-09	2003-09	2003-09
AIC	932	808	775	920	797	764
BIC	960	834	800	951	826	793
Ave. obs. per group	5.3	5.0	5.4	5.3	5.0	5.4
Num. of countries	30	27	23	30	27	23
Observations	158	136	124	158	136	124

Note: Standard errors are in parentheses. Siginificance is indicated as * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

While calculating the country level indicator for the ownership structure, the number of firms at a country-year data point can be very low. More precisely, if there are just two firms in different ownership categories, having one missing observation for a data point changes the evaluation totally. Therefore, we replicate our baseline regressions by setting a minimum number of observations to calculate ownership structure at each country-year data point equal to 20, 40, and 60. Table 4.3 presents the results. From this table, it can be seen that these restrictions substantially decrease the already low number of observations subject to our country level analysis. Despite this, the results confirm our findings in the baseline regressions, providing a strong support for our hypothesis.

Sensitivity Analysis on the Categorization of Firms

In our baseline estimations, we take category A and B firms as managerial in order to calculate the country level ownership indicator. Another possibility is to take only category A firms as managerial. However, the number of observations in category A is substantially smaller compared to the other categories, and this makes it more likely to have the problem of insufficient sample size in constructing the ownership indicator, which we described in the previous subsection. Therefore, in our sensitivity analysis, we also control for the minimum number of firms to calculate the ownership indicator. Our results are presented in Table 4.4.

The upper panel presents the results from the PQML estimations. In the first column, we replicate our full specification. Although the coefficients of the ownership structure and the interaction term are insignificant, the signs of the coefficients are fully in line with the baseline estimations. However, in the next columns, as we increase the minimum number of firms for each country-year data point, the insignificance of these terms dissappears, confirming our baseline results. Again, we see the importance of controlling for the minimum required number of firms in calculating the ownership indicator. In the lower panel, we present the results of using the negative binomial strategy. The results exhibit a similar pattern to the PQML estimations, such that the p-values (not reported) decrease substantially as we increase the minimum number of firms for each data-point. However, this time, the interaction term never becomes significant.

Table 4.4: Taking Only Category-A Firms as Managerial

Dependent variable: Patent counts

	Poisson QML			
Independent variables	(1)	(2)	(3)	(4)
Fuel price	-2.280***	-2.433***	-2.558***	-2.556***
$\ln(1+P_{i,t-1})$	(0.521)	(0.507)	(0.476)	(0.476)
Ownership structure	-0.239	-0.341	-0.502*	-0.474*
$Ownstr_{i,t-1}$	(0.276)	(0.273)	(0.285)	(0.273)
Interaction	0.372	0.511	0.893*	0.847*
$\ln(1+P_{i,t-1}) * Ownstr_{i,t-1}$	(0.488)	(0.494)	(0.529)	(0.508)
Firm num. per data point	Full	>20	>40	>60
Years	2003-09	2003-09	2003-09	2003-09
AIC	1077	948	828	795
BIC	1106	975	854	821
Ave. obs. per group	5.5	5.3	5.0	5.4
Num. of countries	34	30	27	23
Observations	186	158	136	124
	Negative Binomial			
Fuel price	-2.124***	-2.393***	-2.651***	-2.665***
$\ln(1+P_{i,t-1})$	(0.477)	(0.473)	(0.477)	(0.482)
Ownership structure	-0.023	-0.199	-0.496	-0.452
$Ownstr_{i,t-1}$	(0.404)	(0.420)	(0.420)	(0.425)
Interaction	-0.014	0.224	0.932	0.855
$\ln(1+P_{i,t-1}) * Ownstr_{i,t-1}$	(0.623)	(0.654)	(0.694)	(0.704)
Firm num. per data point	Full	>20	>40	>60
Years	2003-09	2003-09	2003-09	2003-09
AIC	1058	929	811	777
BIC	1090	960	840	806
Ave. obs. per group	5.5	5.3	5.0	5.4
Num. of countries	34	30	27	23
Observations	186	158	136	124

Table 4.5: R&D Subsidies

Dependent variable: Patent counts				
	PQML	NBML	PQML	NBML
Independent variables	(1)	(2)	(3)	(4)
R&D bud. appropriations	0.190	0.245**		
$RDapp_{i,t-1}$	(0.160)	(0.106)		
R&D expenditure			-0.001	-0.001
$ln(RDexp)_{i,t-1}$			(0.001)	(0.001)
Ownership structure	0.970***	1.001**	0.239	0.109
$Ownstr_{i,t-1}$	(0.373)	(0.439)	(0.254)	(0.230)
Interaction	-0.380**	-0.405*	0.001	0.001
$Ownstr_{i,t-1} * ln(RD)_{i,t-1}$	(0.184)	(0.219)	(0.001)	(0.001)
ln(GDP per capita)	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes
Years	2003-09	2003-09	2003-09	2003-09
AIC	935	909	400	394
BIC	963	941	422	419
Ave. obs. per group	5.5	5.5	4.6	4.6
Num. of countries	33	33	19	19
Observations	181	181	88	88

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Our choice of functional transformation for a variable is based on the Akaike and Bayesian Information Criterion (AIC and BIC).

R&D Subsidies

In Table 4.5, instead of using tax-inclusive fuel prices and taxes which indicate stricter environmental constraints on firms, we use proxies for R&D subsidies for environmental concerns which relax the environmental constraints on firms. More specifically, we use the fraction of government R&D budget appropriation and expenditure in total government expenditure on environmental issues. In columns (1) and (2), we present the results of the estimations with budget appropriations. In column (1), we present the PQML estimations. The direct effect of subsidy is insignificant, but with a positive sign, confirming the negative effect of tax-inclusive fuel prices. The interaction of subsidy with ownership structure is significant, indicating that relaxing the environmental constraints leads to a differential behavior depending on the ownership structure. The sign of this coefficient is negative, in line with the positive sign of the interaction term in our baseline regressions, where we use tax-inclusive fuel prices. This result is robust to using the negative binomial regression which is presented in column (2). Moreover, the direct effect of budget appropriation is now significant with a positive sign. Akaike and Bayesian information criteria favor the NBML estimations, which is fully in line with our predictions.

In columns (3) and (4), we use the expenditure data. As one can see in Table 4.5, the number of observations is only 88 in these regressions, due to the limited size of the expenditure data. Clearly, it is hard to expect significant results. Columns (3) and (4) confirm this expected result. All the coefficients are highly insignificant.

In these regressions, the only result conflicting with the baseline regressions is the highly significant and positive sign of the ownership structure in columns (1) and (2). According to our model, this is possible only if the fixed costs of R&D is sufficiently higher for the managerial firms compared to the owner-controlled firms. Note, our theoretical predictions are independent of the relative innovation size of these two types of firms.

Sample Selection Problem in Constructing the Country-level Ownership Indicator

Our country level ownership indicator, the fraction of managerial firms, is calculated by using a sample of firms with a known ownership status. The source of the "unknown" ownership status is potentially related to the size of firms, since small firms are not

Table 4.6: Known Ownership Status and Size

Dependent variable: Ownership status (= 1, *if known*; 0, *otherwise*)			
Size category of firms	(1)	(2)	(3)
Very large	1.745***	1.187***	0.697***
(β_1)	(0.236)	(0.218)	(0.189)
Large	1.048***	0.490***	
(β_2)	(0.149)	(0.116)	
Medium	0.558***		-0.490***
(β_3)	(0.101)		(0.116)
Small		-0.558***	-1.048***
(β_4)		(0.101)	(0.149)

Note:Below the coefficients, standard errors are presented in parenthesis.All estimations employ firm specific fixed effects and a full set of time dummies.

generally obliged to report their ownership information. If this is the case, there is a potential sample selection problem in the construction of the ownership indicator, which arises, if size of firms is also a determinant of ownership structure. In this section, we control for this potential problem.

In order to reveal this correlation more clearly, by using all the listed active firms[35], we conduct a fixed effect logit estimation where ownership status, which is a dummy variable indicating known ownership status with 1, regressed on a set of dummies indicating the size category of firms. Results in Table 4.6 indicate that the probability of having a known ownership status is significantly higher for larger firms. These results confirm our expectation that larger firms are more likely to have a known ownership status. As a result, our sample of firms with a known ownership status is not fully representative in terms of size. If firm size and ownership concentration are also correlated, this may constitute a sample selection problem in constructing the country level

[35]We use the term “listed” in order to indicate that a firm is listed at least one of the updates of ORBIS databes. For a detailed explanation see appendix A.4, where we also extensively explain the source of the missing observations in our ownership data.

Table 4.7: Ownership Category and Size

Dependent variable: Ownership status (= 1, $if\ A\&B$; $0\ if\ C\&D$)	(1)	(2)	(3)
Very large firms	-0.393	0.384	-0.022
(β_1)	(0.436)	(0.395)	(0.302)
Large firms	-0.371	0.406	
(β_2)	(0.329)	(0.275)	
Medium sized firms	-0.777***		-0.406
(β_3)	(0.194)		(0.275)
Small firms		0.777***	0.371
(β_4)		(0.194)	(0.329)

Note:Below the coefficients, standard errors are presented in parenthesis. All estimations employ firm specific fixed effects and a full set of time dummies.

ownership indicator.[36]

In order to see if there is a correlation between ownership and size category, we use our observations with a known ownership status. In Table 4.7, the results from a fixed effect logit model are presented. In contrast to the arguments in the corporate governance literature, our results do not indicate a systematic relationship between ownership structure and size of firms. The only problematic situation arises between the medium and small sized firms. This anomalous finding may be due to a situation where small firms are more inclined to report their ownership information when the number of dominant owners is higher.

Irrespective of the reason for the problem revealed by Table 4.7, we can simply control for it by excluding small firms from our baseline regressions. Table 4.8 presents the results from replication of our baseline regressions excluding the small firms. The results are fully in line with the baseline estimations. Therefore, our results are not likely to be driven by a sample selection problem.

[36]However, the problem arises only if the sample selection pattern is not similar across countries. Otherwise, our calculated ownership indicator will still capture the variation across countries.

Table 4.8: Regressions by Excluding Small Firms

Dependent variable: Patent counts						
	Poisson QML			Negative Binomial		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Fuel price	-2.210***	-2.051***	-2.461***	-2.175***	-2.002***	-2.443***
$\ln(1+P_{i,t-1})$	(0.493)	(0.514)	(0.360)	(0.406)	(0.444)	(0.426)
Ownership structure		0.073	-0.625***		0.073	-0.621**
$Ownstr_{i,t-1}$		(0.086)	(0.204)		(0.094)	(0.266)
Interaction			1.096***			1.044***
$\ln(1+P_{i,t-1}) * Ownstr_{i,t-1}$			(0.389)			(0.375)
ln(GDP per capita)	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
Years	2003-09	2003-09	2003-09	2003-09	2003-09	2003-09
Number of countries	34	34	34	34	34	34
Number of observations	194	186	186	194	186	186
Ave. num.of obs. per group	5.7	5.5	5.5	5.7	5.5	5.5
AIC	1119	1075	1063	1099	1056	1051
BIC	1142	1101	1093	1125	1085	1083

Note: Standard errors are presented in parenthesis. Significance is indicated with *** at 1% , ** at 5%, and * at 10% level.

As argued by Demsetz and Villalonga (2001), ownership structure as a possible factor determining firm performance is possibly endogenous. Our analysis in this section also controls for such a concern. Although we control for firm size, there are other factors of relevance to ownership structure which could profitably be included in our ownership determination regression. However, data is not available or relevant for these factors for most of our observations, since these factors are mostly related to and reported by publicly listed firms, which constitutes a small portion of our sample of firms.

4.5. Conclusion

In this paper, we provide both theoretical and empirical evidence in support of our hypothesis that, when ownership structure in a country evolves towards a phase of ownership-managership separation, more stringent environmental regulations become more innovation friendly. Our empirical results show that the negative direct effect of fuel prices on innovation is substantially mitigated through its indirect effect in relation to the ownership structure. However, we find no evidence for a full compensation. The results are robust to many considerations, such as a sample selection problem in constructing the ownership indicator.

We approach the problem from a macro perspective. This is a natural choice from several points of view. Firstly, our policy variable reflects a country level variation. Secondly, we are interested in the effect of our policy variable on overall innovation. Thirdly, country level data is easily accessible through several databases. The price of using country level data is that we have a small number of observations. Despite this disadvantage, our empirical evidence provides significant support to our hypothesis.

On the other hand, it is also necessary to complement our study with a firm level study. Obviously, this approach can also account for the problems that may arise during the aggregation processes. However, a firm level study should account for the fact that the policy variables are country specific. Secondly, merging the firm level patent and the ownership data from different databases is a substantially time consuming process. In order to deal with this problem, future research can concentrate on publicly traded firms, for which such a dataset is more accessible. This strategy will also allow to use citations to weight patent counts, since ownership data for publicly traded firms can be found for a longer time horizon.

4.A. Appendix

4.A.1. Jackknife Method to Calculate Standard Errors

Table 4.9: Negative Binomial Estimations with Jackknife Standard Errors

Dependent variable: Patent counts				
Independent variables	(1)	(2)	(3)	(4)
Fuel price	-2.344***	-2.456***	-2.493***	-2.500***
$\ln(1 + P_{i,t-1})$	(0.672)	(0.683)	(0.709)	(0.688)
Ownership structure	-0.606	-0.667	-0.682**	-0.659**
$Ownstr_{i,t-1}$	(0.532)	(0.522)	(0.274)	(0.269)
Interaction	1.011	1.152	1.377**	1.343**
$\ln(1 + P_{i,t-1}) * Ownstr_{i,t-1}$	(0.699)	(0.728)	(0.556)	(0.538)
ln(GDP per capita)	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes
Firm num. per data point	Full	>20	>40	>60
Years	2003-09	2003-09	2003-09	2003-09
AIC	1050	920	797	764
BIC	1082	951	826	793
Ave. obs. per group	5.5	5.3	5.0	5.4
Num. of countries	34	30	27	23
Observations	186	158	136	124

Note: Standard errors are in parentheses. Significance is indicated with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Although all our baseline regressions control for overdispersion, standard errors of the negative binomial estimations are still sensitive to a misspecification of the conditional variance. In order to account for this potential problem, we replicate our regressions

Table 4.10: Using Industry Diesel Fuel Prices

Dependent variable: Patent counts

	Poisson QML			Negative Binomial		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Fuel price	-2.276***	-2.048***	-2.070***	-2.339***	-2.069***	-2.158***
$\ln(1 + P_{i,t-1})$	(0.411)	(0.419)	(0.340)	(0.375)	(0.420)	(0.402)
Ownership structure		0.078	-0.504**		0.096	-0.501*
$Ownstr_{i,t-1}$		(0.083)	(0.196)		(0.097)	(0.256)
Interaction			1.098**			1.061**
$\ln(1 + P_{i,t-1}) * Ownstr_{i,t-1}$			(0.475)			(0.422)
ln(GDP per capita)	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
Years	2003-09	2003-09	2003-09	2003-09	2003-09	2003-09
AIC	1029	1009	1000	1011	991	987
BIC	1052	1034	1028	1036	1020	1019
Ave. obs. per group	5.9	5.7	5.7	5.9	5.7	5.7
Num. of countries	31	31	31	31	31	31
Observations	182	178	178	182	178	178

Standard errors are in parentheses. Significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

by using the Jackknife method to calculate the standard errors. Table 4.9 presents the outcomes of applying this. In the first column, we see that there is a substantial deterioration in the significance of the coefficient of the interaction term compared to the baseline estimations. However, when we increase the minimum number of observations to calculate the ownership indicator in the following columns, the effect of the interaction term becomes highly significant, confirming the baseline estimations. We note in Table 4.9 that these restrictions decrease the number of observations substantially, which is expected to yield higher standard errors. The opposite situation reveals the importance of controlling for the problem we described in the previous section.

Table 4.11: Excluding 2009 Data

Dependent variable: Patent counts

	Poisson QML			Negative Binomial		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Fuel price	-1.986***	-1.964***	-2.188***	-2.012***	-1.941***	-2.203***
$\ln(1 + P_{i,t-1})$	(0.378)	(0.342)	(0.346)	(0.406)	(0.456)	(0.455)
Ownership structure		0.003	-0.536***		0.029	-0.527*
$Ownstr_{i,t-1}$		(0.065)	(0.151)		(0.089)	(0.272)
Interaction			0.849***			0.843**
$\ln(1 + P_{i,t-1}) * Ownstr_{i,t-1}$			(0.233)			(0.392)
ln(GDP per capita)	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
Years	2003-09	2003-09	2003-09	2003-09	2003-09	2003-09
AIC	839	803	799	835	799	796
BIC	857	824	823	856	823	823
Ave. obs. per group	4.8	4.7	4.7	4.8	4.7	4.7
Num. of countries	33	32	32	33	32	32
Observations	159	151	151	159	151	151

Standard errors are in parentheses. Significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.A.2. Using Industry Fuel Price Data

Our preferred fuel price series is "Household diesel fuel prices", since it is the largest among all alternatives. IEA also provides data for industry prices. We present the replication of our baseline regression by using this alternative measure in Table 4.10. The number of observations is now smaller. However, the results are fully in line with the baseline regressions.

4.A.3. Excluding 2009 Data

For our patent data from EUROSTAT, the last year of the dataset is indicated as provisional, since these values may be subject to change in the future, due to lags in the patenting procedure. For this reason, we replicate our baseline regressions by exclud-

Table 4.12: Frequency and Percentage of Firms with (Un)Known Ownership Indicator by Size in 2009

Size	Ownership Indicator			
	Unknown	Known	Not Listed	Total
Unknown	0	0	41926	41926
	0%	0%	100%	100%
Very large	320	1819	0	2286
	15%	85%	0%	100%
Large	2895	3976	0	7031
	42%	58%	0%	100%
Medium	9696	5233	0	15465
	65%	35%	0%	100%
Small	15038	5112	0	20929
	75%	25%	0%	100%
Total	27949	16140	41926	86015
	32%	19%	49%	100%

ing 2009 data. The results, presented in Table 4.11, are fully in line with our baseline regressions.

4.A.4. Source of Missing Ownership Data and Sample Selection Problem

Firstly, we start with some descriptive statistics in order to reveal the source of a potential sample selection problem. Table 4.4 presents the number of firms and their percentages with known/unknown ownership status by size in 2009. The missing ownership indicator can be due to two factors. Firstly, a firm may be listed in 2009 in the ORBIS database with an unknown ownership indicator. We label these observations as "unknown". Secondly, the firm may not be listed for 2009 in any updates of the ORBIS database. In merging the updates of this database, these observations are labeled as "not listed". In total, we have 86,015 firms observed throughout our sample period. These are all legally registered firms in their countries. Firstly, for 2009, 49% of these firms are not listed in any updates, therefore we have no 2009 data for them.[37] Secondly, the percentage of

[37]However, we note that each of these 86,015 firms is listed in at least one year from 2003 to 2009.

firms indicated with an unknown ownership status is 32. In performing our regressions, we also exclude observations where the firms are not active in a given year. Therefore, our ownership indicator for 2009 is calculated by using 16% of the whole population of firms (This is not reported in Table 4.4). This constitutes 13,867 observed firms out of 86,015, which is a reasonable sample size.

The distinction between the data points labeled as “not listed” and “unknown” is crucial to our analysis. Note that all the firms in our sample are listed for at least one year. However, BvD does not collect the data for a specific firm every year, so this firm may have missing observations, which are labeled as “not listed”. Also considering that ownership structure is argued to be a persistent variable at the firm level, these missing observations are not expected to cause a sample selection problem. Another source of missing observation is that a listed firm may not be reporting the ownership information for some reason. These are the missing observations labeled as “unknown”. The source of “unknown” ownership status is likely to be related to size of firms, since smaller firms are not generally obliged to report their ownership information. In Table 4.12, a known ownership status seems to be positively correlated with increasing size.

Table 4.12 also reveals an important opportunity to control for the possible sample selection problem depending on size. It is seen that we have the size category of all the listed firms in 2009.[38] Therefore, following the literature on the determinants of ownership structure, we can use the size data to estimate the ownership indicator of all the listed firms. Moreover, this can increase our sample size to 43% of the whole population.

[38] We note that this situation is also valid for the years following 2004.

Chapter 5

Intra Elite Conflict, Collective Action Problem of the Masses, and Political Transitions

5.1. Introduction

A substantial part of the political economy literature argue that democratic transitions are conscious political choices of the elites for their own benefit; however, there are broadly two perspectives depending on the highlighted force which leads the elite to transfer substantial political power to the masses. Firstly, according to *the elite-poor conflict view*, democratic transitions are the result of a revolutionary threat originating from the masses. As an example, Acemoglu and Robinson (2000) consider an autocratic society with an elite and a poor class, where income allocation is determined by redistributive policies. It is argued that the revolutionary threat by the poor due to high income inequality is the main driving force of the elite-led democratic transition, which, in the end, guarantees future redistribution for the poor. The other approach, seeks the answer in elite structure, following the tradition initiated mainly by Mosca et al. (1939) and Pareto (1991). According to *the elitist view*, the elites are unchecked and lose their privilege only through infighting. Not surprisingly, these two perspectives put forward different preconditions for a democracy to consolidate. According to Acemoglu and Robinson (2001), democracy consolidates only if the society is sufficiently equal, alleviating the class conflict over the redistributive policy. On the other hand, Higley and Burton (1989) argue that disunity in the elite structure, originating from nation-state formation, is the factor leading to unstable regimes.

Despite the substantial literature from these two perspectives, the relation between the intra-elite and the elite-poor conflict is a relatively uninvestigated area. Are they two competing perspectives, do they emerge in different circumstances, or is there a degree of complementarity between these factors leading to democratic transitions? In this paper, we put forward a potential link between the roles of intra-elite and elite-poor conflict in democratic transitions, and draw conclusions about the possible consequences for regime stability. At the center of our analysis, there is the collective action problem of the masses and intra-elite conflict, forcing some elite factions to employ potential de facto power of the masses.

There are two main difficulties in incorporating the elite structure (unity vs. disunity) as a possible determinant of political regimes. Firstly, elite structure varies substantially across societies and over time (Mosca et al., 1939). Naturally, the focus of the early studies was to identify some major forms of elite structure, depending on the character and the degree of the contest within the elite, which lead to distinctive political regimes. Although there is a loose agreement on this issue in the political science literature, one widely accepted idea is that a unified elite structure is a precondition for a stable regime. However, the problem of formalizing the dispute within the elite and explaining how a consensus arises, paving the way to a stable regime, remains very dependent on the country and time period under consideration. Secondly, a further complication arises, if one tries to establish a casual link between the elite structure and democratic transitions. If we define democratic transition as a significant transfer of political power to the masses (Acemoglu and Robinson, 2000), it seems more appropriate to handle the problem as an elite-poor conflict. The question is: why do the elites need democracy as a solution to their internal conflict, while the main beneficiary is a third party, the poor? Ghosal and Proto (2009) argue that democracy is a result of intra-elite conflict as a solution to uncertainty in future political power allocation. An earlier approach argues that the elites extend franchise as a strategic choice by expecting the support of newly enfranchised (Collier, 1999). Both approaches assume that the non-elite have no political power in non-democracy, due to the high costs of collective action imposed by the elite. Therefore, democracy is seen as an institution to regulate disputes within the elite. Giving no role to the masses in the democratization process, the elitist view can be criticized for being unable to explain elite unification and democratic transitions within the same framework. The reason is that once we have a unified elite, we would not observe a democratic transition according to the above explanations. So, elite unification remains as a precondition for a stable regime, but has no consequences for regime transitions. In order to identify the role of the elite structure in democratic transitions, our strategy in this paper is to relate the intra-elite conflict to the elite-poor conflict.

Rather than identifying the sources of the contest and the strategies which the elites adopt against each other, which is problematic due to the variability of elite structures, we identify when the intra-elite conflict arises as a threat to regime in non-democracies. We argue that in any non-democracy, there is an inherent conflict over income and political power distribution within the elite. The disunity of the elite is persistent; however, only if an elite faction, discontented with the existing regime, is willing and able to mobilize the masses against the authority, an intra-elite conflict arises as a threat to the regime. Therefore, in our model, a unified elite is characterized by the inability of the discontent elite factions to mobilize the masses against the regime.

The critical element in our analysis is the collective action problem of the poor, which establishes the link between the intra-elite and the elite-poor conflict. In our model, democratic transition can take place due to a revolutionary threat by the poor, even if the poor are not capable of overcoming collective action problems with their own resources. The underlying idea is that maintaining a collective movement is a costly activity, and when the poor alone cannot finance the pre-revolutionary activity, intra-elite conflict may force an elite faction, which is discontent with the existing regime, to bear the costs of the collective action in order to employ de facto political power of the poor. In such an environment, the revolutionary threat cannot be credible unless there is an appropriate level of conflict among the elite, which is assumed to depend on intra-elite income inequality. Therefore, our model distinguishes between two types of democratization. One is due to a revolutionary threat posed by the unconstrained poor in solving the collective action problem, which we call the "poor-elite conflict transition". The second type is due to a revolutionary threat posed by a revolutionary coalition between the resource-constrained poor in maintaining collective action and a discontent elite faction, which we call "intra-elite conflict transition". Such a setting allows us to investigate when the elite-poor conflict or the intra-elite conflict becomes the major factor in democratic transitions, and why.

Our results indicate that intra-elite conflict transition does not arise in relatively equal societies, which is argued to be a factor leading to consolidated democracies (Acemoglu and Robinson, 2000). Thus, elite unification arises when the income distribution is sufficiently equal. Therefore, it is shown that elite unification and low income inequality are not separate, but consistent preconditions for a consolidated democracy. That is, the elitist and the elite-poor conflict perspectives are not conflicting views.

Secondly, our model explains why some mass movements create post revolutionary non-democracies, such as in Russia and China in the first half of the twentieth century, while some revolutions lead to democracies, such as in 19^{th} century France. Our answer is related to ability of discontented elite factions, leading the poor for a revolution, in compensating collective action cost of the masses by making pre-revolution transfers, while keeping post-revolutionary income share of the poor sufficiently low, which, in the end, prevents the masses to act collectively after a revolution. As a result, a post-revolutionary non-democracy arises. We show that such an outcome requires a sufficiently low intra-elite inequality. This can also be interpreted as a second kind of elite unification, which is consistent with the view that elite unification can also lead to consolidated non-democracies (Higley and Burton, 1989).

Relevance of our intra-elite conflict model for revolutions is based on two observations. Firstly, in history, revolutions replaced monarchies with state organizations in line with the interests of emerging elites who pursue an industrial development path. Secondly, every revolution witnessed a substantial mass mobilization organized around the elites who are discontented with the existing institutions. Therefore, the intra-elite and the elite-poor conflicts play interrelated roles in revolutions, as well as in democratic transitions.

5.2. The Model

Our model is an infinitely repeated discounted dynamic game, following the approach by Acemoglu and Robinson (2006), where democracies are argued to be more redistributive compared to non-democracies. We extend their model by introducing elite heterogeneity and allowing for coalition formation between some elite factions and the poor. Initially we are in a non-democracy. Political power is in the hands of the elite class, but its within class distribution is not necessarily symmetric. We call the politically more powerful elite group, the strong elite. They can change future income distribution through redistributive policies. Income levels of any elite group are assumed to be higher than the average income level. Therefore, there is no intra-elite conflict over redistributive policies. The unequal intra-elite income distribution is the only source of the discontent among the elite.[39]

5.2.1. The Environment

Consider a society with three classes of homogenous agents: the strong elite (s), the weak elite (w), and the poor (p). The strong elite and the weak elite constitute a fraction, respectively, λ^s and λ^w of the total population. The population is normalized to one. Thus, $1 - \lambda^s - \lambda^w$ of the society is poor and they are the majority, such that $1 - \lambda^s - \lambda^w = \lambda^p > 1/2$. Initially, each agent is endowed with an income which is denoted with y^i, where $i = s, w, p$ indicates the individual's class. That is, for each class, income is uniform across agents. Denoting the average income with $\bar{y}$, income-share parameters which define the fraction of income accruing to each class are given as follows:

$$\gamma^s = \frac{y^s \lambda^s}{\bar{y}}, \ \ \gamma^w = \frac{y^w \lambda^w}{\bar{y}}, \ \ \gamma^p = \frac{y^p \lambda^p}{\bar{y}}, \ where\, \gamma^s + \gamma^w + \gamma^p = 1.$$

Agents' income can be written in terms of income-share parameters as follows:

$$y^s = \frac{\gamma^s \bar{y}}{\lambda_s}, \ \ y^w = \frac{\gamma^w \bar{y}}{\lambda^w}, \ \ y^p = \frac{(1 - \gamma^s - \gamma^w)\bar{y}}{1 - \lambda^s - \lambda^w}.$$

The income-share parameters are convenient for formalizing and interpreting our model. However, analyzing the comparative statics with inequality parameters makes more sense. Firstly, the elite-poor inequality is defined as the fraction of the total income owned by the elite. Secondly, the intra-elite inequality is defined as the fraction of elite income owned by the strong elite. Elite-poor inequality (κ_1) and intra-elite inequality (κ_2) are parametrized as follows:

[39] In order to explain the gradual expansion of enfranchisement, Acemoglu and Robinson (2006) model the intra-elite conflict as an occurrence between the rich and the middle class whose income is less than the average. So the middle class has conflicting interests to the rich over the redistributive policy. However, in our model we do not consider a middle class, which may have substantial political power in a democracy.

$$\kappa_1 = \frac{y^s\lambda^s + y^w\lambda^w}{\bar{y}}, \;\; \kappa_2 = \frac{y^s\lambda^s}{y^s\lambda^s + y^w\lambda^w}.$$

We can also express agents' income in terms of the inequality parameters:

$$y^s = \frac{\kappa_1\kappa_2\bar{y}}{\lambda^s}, \;\; y^w = \frac{\kappa_1(1-\kappa_2)\bar{y}}{\lambda^w}, \;\; y^p = \frac{(1-\kappa_1)\bar{y}}{1-\lambda^s-\lambda^w}.$$

Some regularity conditions follow. Firstly, since it is assumed that $1-\lambda^s-\lambda^w > 1/2$, we have $\lambda^s + \lambda^w < 1/2$. Secondly, imposing the assumption that $y^p < \bar{y}$, $y^w > \bar{y}$ and $y^s > \bar{y}$ leads to the following conditions:

$$\frac{1-\gamma^s-\gamma^w}{1-\lambda^s-\lambda^w} < 1, \;\; 1 < \frac{\gamma^w}{\lambda^w}, \;\; 1 < \frac{\gamma^s}{\lambda^s}. \tag{5.1}$$

Or in terms of the inequality parameters, we have

$$\frac{1-\kappa_1}{1-\lambda^s-\lambda^w} < 1, \;\; 1 < \frac{\kappa_1(1-\kappa_2)}{\lambda^w}, \;\; 1 < \frac{\kappa_1\kappa_2}{\lambda^s}. \tag{5.2}$$

Cost of Taxation, Government Budget Constraint, and Most Preferred Tax Rates

Income allocation is determined by the class with dominant political power. Any class which holds de jure political power can impose a tax rate, $0 < \tau < 1$, on income. In terms of de facto political power, the poor can impose a revolutionary threat either by itself, or by forming a coalition with the weak elite against the strong elite in order to change the income distribution.

We assume that taxation is costly. The cost of taxation, $C : [0,1] \to R_+$, is a fraction lost of total income. It is assumed to be a function of the tax rate. In order to ensure that the second order conditions of the agents' utility maximization problem will be satisfied, we assume $C'(\tau) > 0$ and $C''(\tau) > 0$. Moreover, in order ensure an interior solution, we assume $C^{'}(0) = 0$ and $C^{'}(1) = 1$. The government collects income taxes, and redistributes it uniformly to all citizens. Denoting the transfer with T, the government budget constraint can be written as $T = \sum_{j=1}^{n} \tau y^j - C(\tau)n\bar{y}$. Here, n stands for the population size and agents are indexed with i. Normalizing the population to one leads to $T = (\tau - C(\tau))\bar{y}$.

In every period, each agent earns the income described previously. The intertemporal expected utility of every agent at time $t = 0$ is the discounted sum of post-tax incomes which can be stated as follows:

$$V^i = E_0 \sum_{t=0}^{\infty} \beta^t \hat{y}_t^i,$$

where $\beta \in (0,1)$ is the discount factor, and $\hat{y}_t^i$ stands for the post-tax income which is totally channeled to consumption. So, the indirect utility of an agent from class i, in case of no regime transitions, can be formulated as:

$$V(y^i \mid \tau) = \frac{(1-\tau)y^i + (\tau - C(\tau))\bar{y}}{1-\beta}.$$

Maximization of the indirect utility with respect to the tax rate gives the following first order condition:

$$\begin{aligned} -y^i + (1 - C^{'}(\tau))\bar{y} &= 0 \, and \, \tau^i > 0 \, or \\ -y^i + (1 - C^{'}(\tau))\bar{y} &< 0 \, and \, \tau^i = 0. \end{aligned} \tag{5.3}$$

Under the assumptions on the cost function, if individual's income is higher than average income, we are in the second line of condition (5.3). That is, under the assumed redistributive policy, any agent above the average income is worse off with a positive tax rate. Therefore, for any elite, the ideal tax rate is zero. On the other hand, for the poor, the first line of condition (5.3) applies, and they prefer a positive tax rate. Applying the implicit function theorem to the first order condition of the poor, we obtain the effect of income on the ideal tax rate of the poor:

$$\tau^{'}(y^p) = -\frac{1}{C^{''}(\tau(y^p))} < 0. \tag{5.4}$$

Therefore, the poor prefer higher tax rates as they are further away from the average income. The most preferred tax rate for each class is given as follows:

$$\tau^s = \tau^w = 0 \tag{5.5}$$

$$C^{'}(\tau^p) = 1 - \frac{\gamma^p}{\lambda^p}. \tag{5.6}$$

Equation (5.6) follows from the first order condition of the poor and the definition of income-share parameters. Both sides of the equation (5.6) are positive from the regularity conditions given by equation (5.1) and the assumptions imposed on the cost function. Equation (5.6) can be written in terms of the inequality parameters as follows:

$$C^{'}(\tau^{p}) = 1 - \frac{1-\kappa_{1}}{\lambda^{p}}\,. \tag{5.7}$$

Here, the term on the right hand side is positive by the regularity condition given by equation (5.2).

From equation (5.6), it can be verified that, as the per capita income share of the poor increases, they demand less redistribution. In the same way, we can derive the effect of the inequality parameters from equation (5.7). Taking derivatives of both sides of (5.7) with respect to κ^{1} gives $\partial\tau^{p}/\partial\kappa^{1} = 1/(\lambda^{p}C^{''}(\tau^{p}))$, which is positive due to the assumptions made on the cost function. Therefore, higher elite-poor inequality leads the poor to demand higher redistribution.

Revolution

Due to an unequal income distribution, the poor can initiate a revolution which is assumed to be always successful. However, a revolution is attempted only if the poor can afford the cost of collective action, denoted with ε, which is assumed to be a fraction of total income. If the poor is resource-constrained in solving the collective action problem, another possibility is that the weak elite and the poor can form a revolutionary coalition. More precisely, the weak elite finances the uprising by transferring Ω fraction of their income to the poor, and benefits from the post-revolution income allocation. If the revolution takes place, all post-revolutionary income is allocated to agents of the revolting classes, and the strong elite ends up with zero indirect utility. The post-revolution income share for coalition members is determined prior to the revolution. In order to ease interpretation, we maintain a simple structure for pre-revolution agreements on post-revolution income allocation. We assume that the weak elite offers a fraction θ of post-revolution income to the poor, and the poor decides whether to revolt by observing the offered combination of Ω and θ.[40]

[40]One might expect a commitment problem relating to post-revolutionary income allocation. However, note that a potential commitment problem in post-revolutionary income allocation and future redistribution in non-democracy are very different. Firstly, post-revolutionary income allocation, which takes place in the same period with pre-revolutionary agreements, is a one period event, as opposed to

The cost of collective action, at the date when a revolution takes place, is denoted by ε^s, where $s = \{L, H\}$. We assume that it fluctuates between the values ε^H and ε^L. It is assumed that $\varepsilon^L = 1$, which stands for a state where the cost of collective action does not allow a revolution. This is simply a normalization. On the other hand, $\varepsilon^H = \varepsilon$, which is smaller than one. It is assumed that $\varepsilon_t = \varepsilon^H$ with probability q, independent of ε_{t-1}. Therefore, the threat of revolution fluctuates over time, indicating the transitory nature of de facto political power held by the potential revolters. Hence, the strong elite may change the policy to prevent revolution in every period, in response to the level of de facto power held by the poor.

Revolution also leads to post-revolutionary costs. Some fraction μ of total income is destroyed forever during the revolution. It is assumed that $\mu < 1$, otherwise revolutionary threat would never be credible. This makes sense by considering that there should exist some assets that would never be totally destroyed in a revolution, such as human capital.

Firstly, we start by formalizing the instantaneous payoffs when the poor can revolt alone, and expropriate the income of both the weak and the strong elite, which leads to zero indirect utility for both elites. The poor contributes $\varepsilon\bar{y}$ to overcome the collective action problem. Once the revolution occurs, the income of the strong and the weak elites is allocated uniformly to the poor agents, which is $(\gamma^s + \gamma^w)\bar{y}/\lambda^p$. However, during the revolution $1 - \mu$ of the total income is destroyed forever. Therefore, each poor agent gets $(1 - \mu)[\bar{y}/\lambda^p - \varepsilon\bar{y}]$ following a revolution. Also, the poor face the cost of collective action only once, so in each following period the poor get $(1 - \mu)\bar{y}/\lambda^p$. Rearranging these payoffs, the intertemporal indirect utilities of each agent following a revolt by the poor alone is given by:

redistribution, which are taken into consideration for an infinite horizon. Secondly, we assume that what is allocated between the parties is the productive resources of the economy. Therefore, one expects the coalition members to foresee sufficiently accurately how much of the productive resources they can own following a revolution. A final point is that once the productive resources are allocated, it is also more binding for future income allocation compared to a promise for future redistribution. Indeed, redistributions inherently lead to commitment problems, since the ownership of productive resources remains unchanged. For these reasons, we assume that coalition members know the fraction of post-revolutionary income which they will be able to expropriate, if they undertake a revolution.

$$\begin{aligned} V^s_{pe}(R,\mu) &= V^w_{pe}(R,\mu) = 0 \\ V^p_{pe}(R,\mu) &= \frac{(1-\mu)\bar{y}}{\lambda^p(1-\beta)} - (1-\mu)\varepsilon\bar{y}, \end{aligned}$$

where the subscript "*pe*" indicates the case where the poor-elite conflict arises. Secondly, the payoffs are different if the poor revolt by forming a coalition with the weak elite. In this case each agent of the weak elite contributes a fraction Ω of their income to solve the collective action problem of the poor, and, in turn, gets the $1-\theta$ fraction of the remaining income following the revolution. Therefore, the payoffs are given as follows:

$$\begin{aligned} V^s_{ie}(R,\mu) &= 0 \\ V^w_{ie}(R,\mu) &= (1-\mu)[\frac{(1-\theta)\bar{y}}{\lambda^p(1-\beta)} - \Omega y^w] \\ V^p_{ie}(R,\mu) &= (1-\mu)[\frac{\theta\bar{y}}{\lambda^p(1-\beta)} - \varepsilon\bar{y} + \frac{\Omega y^w \lambda^w}{\lambda^p}], \end{aligned}$$

where the subscript "*ie*" indicates the intra-elite conflict case.

Expected Payoffs of Not Revolting

In non-democracy, the strong elite can prevent the revolution via redistribution. However, it is not certain that the redistribution will take place in the future. Therefore, all agents discount their indirect utility in case of no revolution with a probability reflecting the uncertainty of future implementation of the promised policy.

In state $\varepsilon_t = \varepsilon^L$, there is no threat of revolution, therefore the strong elite will set their most preferred tax rate, which is zero. Then the indirect utility to each agent in state ε^L can be written as follows:

$$V^i(N,\varepsilon^L) = y^i + \beta[qV^i(N,\varepsilon^H) + (1-q)V^i(N,\varepsilon^L)], \tag{5.8}$$

where, in the current period, agents get no redistribution, but in the future the threat state will be ε^H with probability q, and there may be a redistribution.

When the threat of revolution is realized in state $\varepsilon_t = \varepsilon^H$, the strong elite can set a tax rate τ^N to prevent the revolution. However, this cannot be credible for future redistributions due to the transitory nature of de facto power held by the poor. That is, when the state is $\varepsilon_t = \varepsilon^L$ in the future, the strong elite will set their most preferred tax rate, $\tau^N = 0$. Assuming that a redistribution prevents revolution, the intertemporal indirect utility of each agent in state $\varepsilon_t = \varepsilon^H$ can be written as follows:

$$\begin{aligned} V^i(N, \varepsilon^H, \tau^N) &= [y^i + \tau^N(\bar{y} - y^i) - C(\tau^N)\bar{y}] \\ &\quad +\beta[qV^i(N, \varepsilon^H, \tau^N) + (1-q)V^i(N, \varepsilon^L)] \end{aligned} \tag{5.9}$$

Equations (5.8) and (5.9) are two equations with two unknowns. We can solve for $V^i(N, \varepsilon^H, \tau^N)$, which gives:

$$V^i(N, \varepsilon^H, \tau^N) = \frac{y^i + (1-\beta(1-q))(\tau^N(\bar{y} - y^i) - C(\tau^N)\bar{y})}{1-\beta}. \tag{5.10}$$

Due to the commitment problem relating to future redistributions, current redistributions may not suffice to prevent a revolution. Democratization is the instrument for the strong elite to make a credible commitment for future redistributions. Therefore, whenever a redistribution is not sufficient to prevent revolution, the strong elite may democratize to create a higher incentive for the poor not to revolt. In a democracy all political power is in the hands of the poor, since it is assumed that the median voter is poor. Therefore, the tax rate is set by the poor as τ^p. Hence, the payoffs in democracy for an agent of class i is

$$V^i(D) = (y^i + \tau^p(\bar{y} - y^i) - C(\tau^p)\bar{y}))/(1-\beta). \tag{5.11}$$

Timing of Events

Initially we are in a non-democracy where the strong elite holds de jure political power. The timing of events at an arbitrary period t is as follows:

1. The state $\varepsilon_t \in \{\varepsilon^L, \varepsilon^H\}$ is revealed.

2. The strong elite decides whether to democratize, $\phi \in \{0, 1\}$.

3. If $\phi = 1$, the tax rate is determined by the median voter. If $\phi = 0$, the strong elite sets a tax rate.

4. Realizing the tax rate set by the strong elite, the weak elite decides whether to make a transfer to the poor, $\Omega \in [0,1]$. If $\Omega > 0$, they propose a post-revolution income share, $\theta \in [0,1]$, to the poor.

5. The poor decides whether to form a revolutionary coalition with the weak elite.

6. The poor decides whether to revolt, $\rho \in \{0,1\}$.

7. Incomes are realized and consumption takes place.

Class Structure

Our model imposes some structure on the instruments available to the agents. At this point, it is worthwhile to discuss these assumptions briefly. Firstly, our model imposes elite disunity exogenously as an inherent character of societies. Such an assumption is realistic, only if there is a persistency in the elite structure. Higley and Burton (1989) argue that the more common elite structure is characterized by disunity, and once an elite structure emerges, it persists. Indeed, our model adopts the same approach, and we introduce elite disunity to our model exogenously. However, our model explains when it does not play a decisive role in regime transitions. We interpret such a case as elite unity. In this case, only elite-poor conflict is relevant for regime transitions. In other words, our goal is not to explain how such a class structure has evolved in the pre-transition era. This implies that the unit of analysis for our model is a regime transition, whether it is successful or not. For example, the 17^{th}-18^{th} century Britain is characterized by elite disunity, whereas 19^{th} century Britain is characterized by elite unity (Lachmann, 1990). Therefore, these periods are different units to be analyzed.

Secondly, we restrict the intruments available to the strong elite with income redistribution and democratization. It can be assumed that the strong elite also has control of the military, which provides the means to preserve their preferred income allocation. Such an assumption does not change our main results. One can refer to Acemoglu and Robinson (2006) to see the implications of such an assumption.

Thirdly, we allow a coalition only between the weak elite and the poor, since it is the only historically relevant case for most of the regime transitions where the within elite struggle is between a resource owner elite class (such as aristocracy), and an emerging elite class pursuing industrialization. The most important element leading to a coalition with the emerging elites and the poor instead of other possibilities is that emerging elites can commit to post-revolutionary allocation of a productive resource which is land, while aristocracy cannot. This is also the reason why we exclude any commitment problem about the post-revolutionary income allocation within the coalition. However, if we would forcefully allow a coalition also between the strong elite and the poor, the poor would always prefer a coalition with the weak elite, since the poor would always get higher payoff by expropriating the strong elite rather than the weak elite. Note that, although the direct transfers from the strong elite to the poor would be possibly higher, we would always have a parameter space where it cannot compensate the forgone gains of the poor from expropriating the strong elite. Finally, there is no way for the strong and the weak elites to compromise with side-payments, since they always have an incentive to expropriate eachother, as long as they hold some assets. In our setting, this is only possible by getting support of non-elites.

5.2.2. Equilibrium

In order to have a simple recursive structure, we restrict attention to Markow perfect equilibria. That is, we assume that strategies of players at an arbitrary date only depend on the current state of the game. Firstly, our main state variable is the cost of collective action, $\varepsilon = \{\varepsilon^L, \varepsilon^H\}$. Secondly, the political state is denoted by $P = \{N, D, R\}$, where N is non-democracy, D is democracy, and R is revolution. Let σ^i be the strategy space for an agent of class $i = \{s, w, p\}$. The strong elite decides whether to democratize, $\phi : \{\varepsilon^L, \varepsilon^H\} \to \{0, 1\}$, when $P = N$, and sets a tax rate $\tau^N : \{\varepsilon^L, \varepsilon^H\} \to [0, 1]$, if $\phi = 0$. If $\phi = 1$, the state switches to $P = D$. So the strategy space for the strong elite is $\sigma^s(\varepsilon, P) = \{\phi, \tau^N\}$. The weak elite decides on the transfer rate $\Omega : \{\varepsilon^L, \varepsilon^H\} \times \{0, 1\} \times [0, 1] \to \{0, 1\}$, and proposes a post-revolutionary income share $\theta : \{\varepsilon^L, \varepsilon^H\} \times \{0, 1\} \times [0, 1] \to [0, 1]$. That is, their strategy space is $\sigma^w(\varepsilon, P \mid \phi, \tau^N) = \{\Omega, \theta\}$. When $P = N$, the poor decide whether to revolt, $\rho : \{\varepsilon^L, \varepsilon^H\} \times \{0, 1\} \times [0, 1]^3 \to \{0, 1\}$. Secondly, if $P = D$, the poor set a tax rate $\tau^D : \{\varepsilon^L, \varepsilon^H\} \times \{0, 1\} \times [0, 1]^3 \to [0, 1]$. Therefore, the strategy space of the poor is given by $\sigma^p(\varepsilon, P \mid \phi, \tau^N, \Omega, \theta) = \{\rho, \tau^D\}$. Having the strategy spaces, we can define the equilibrium as follows: A Markov perfect equilibrium is a combination of

$$\left\{\sigma^s(\varepsilon, P), \sigma^w(\varepsilon, P \mid \phi, \tau^N), \sigma^p(\varepsilon, P \mid \phi, \tau^N, \Omega, \theta)\right\}$$

such that σ^i is a best response to σ^{-i} for all states.

Collective Action Problem

The poor can afford the cost of collective action, if $y^p - \varepsilon\bar{y} > 0$. In terms of the income share parameters, this condition reads as follows:

$$\varepsilon < \frac{\gamma^p}{\lambda^p}. \tag{5.12}$$

Therefore, condition (5.12) defines a critical ε level for the no coalition case, which is $\varepsilon_1^{pe} \equiv \gamma^p/\lambda^p$. If condition (5.12) does not hold, the weak elite may finance the poor for a revolution. Thus, the resource constraint of the poor becomes $y^p - \varepsilon\bar{y} + \Omega y^w \lambda^w/\lambda^p > 0$. This leads to a lower bound for the transfer rate as $\Omega > (\varepsilon\lambda^p - \gamma^p)/\gamma^p$. We denote the lower bound by $\underline{\Omega}$. However, the transfer is also limited by the budget constraint of the weak elite such that $y^w - \Omega y^w{>}0$, which defines an upper bound for the transfer rate as $\Omega < 1$, denoted by $\overline{\Omega}$. The coalition is feasible only if $\overline{\Omega} > \underline{\Omega}$, which requires

$$\varepsilon < \frac{\gamma^p + \gamma^w}{\lambda^p}. \tag{5.13}$$

This defines another critical level for ε as $\varepsilon_1^{ie} \equiv (\gamma^p + \gamma^w)/\lambda^p$, where the superscript "*ie*" indicates an intra-elite conflict case. These critical levels defines the feasibility of a revolutionary threat by the poor, whether alone, or by in coalition with the weak elite. The situation is described by the following proposition.

Proposition 5.1. *If the cost of collective action to the poor is lower than the critical level defined by inequality (5.12), that is when $\varepsilon < \varepsilon_1^{pe}$, the poor are able to solve the collective action problem with their own resources.*[41] *If $\varepsilon \geq \varepsilon_1^{pe}$, then the poor are resource-constrained, and cannot overcome the collective action problem. In this case, if condition (5.13) holds such that $\varepsilon_1^{ie} > \varepsilon \geq \varepsilon_1^{pe}$, then the poor may prefer to form a coalition with the weak elite. If condition (5.13) is not satisfied, then the collective action problem of the poor cannot be solved, and there can be no revolutionary threat.*

Therefore, the level of cost of collective action with respect to the critical levels is the main factor determining the nature of a potential uprising, whether it is by the poor, or

[41] It will be shown that when the poor are able to solve the collective action problem without a coalition with the weak elite, they always prefer to act alone.

a coalition between the poor and the weak elite. From equations (5.12) and (5.13), the poor's income share should be sufficiently high for a revolution to be feasible. If it is not sufficient for the poor to revolt, then the weak elite's income should be sufficiently high to finance the resource constraint of the poor.

Elite-Poor Conflict Transitions

We start with the case where the poor are able to solve the collective action problem with their own resources. The poor will not be willing to undertake a revolution if the expected utility from the revolution is lower than the expected utility when there is no redistribution forever. Therefore, the revolutionary threat from the poor alone can only be credible, if:

$$V^p_{pe}(R,\mu) > \frac{y^p}{1-\beta}.$$

Otherwise, the strong elite anticipates that the poor will never revolt, and set their ideal tax rate accordingly. By using the income share parameters, the revolution constraint can be written as:

$$\varepsilon < \frac{1}{(1-\mu)\lambda^p(1-\beta)}[1-\mu-\gamma^p]. \tag{5.14}$$

We denote this critical level by ε_2^{pe}. In next sections we will show that if $\varepsilon < \varepsilon_2^{pe}$, our model reduces to Acemoglu and Robinson (2006) model. Following, we restate their results.

In a non-democracy the best concession that the strong elite can offer to prevent revolution is to promise the ideal tax rate of the poor, τ^p, given by equation (5.6). If even this concession cannot prevent the poor from initiating the revolution, then the revolution cannot be prevented by redistribution. We denote the payoff to the poor, when the strong elite sets τ^p, by $V^p(N,\varepsilon^H,\tau^p)$. Substituting $\tau^N = \tau^p$ in equation (5.10), we can find $V^p(N,\varepsilon^H,\tau^p)$. The poor are willing to revolt if $V^p_{pe}(R,\mu) > V^p(N,\varepsilon^H,\tau^p)$. This condition leads us to the following inequality:

$$\varepsilon < \frac{1}{(1-\mu)\lambda^p(1-\beta)}[(1-\mu-\gamma^p)-(1-\beta(1-q))A], \tag{5.15}$$

where $A = (\tau^p(\lambda^p-\gamma^p) - C(\tau^p)\lambda^p)$, which is equal to the fraction of total income transferred to the poor when the tax rate is τ^p. We denote the critical level defined by equation (5.15) by ε_3^{pe}.

If a redistribution cannot prevent revolution, another choice for the strong elite is to democratize the regime, guaranteeing future redistribution for the poor. However, the poor may not find democratization beneficial, if $V_{pe}^p(R, \mu) > V^p(D)$, where $V^p(D)$ is given by equation (5.11). This condition leads to

$$\varepsilon < \frac{1}{(1-\mu)\lambda^p(1-\beta)}[(1-\mu-\gamma^p) - A], \tag{5.16}$$

which defines the critical level ε_4^{pe}.

Now, we summarize the equilibrium for the case where the poor are able to solve the collective action problem with their own resources with the following proposition.

Proposition 5.2. *Assume that the poor are able to overcome the realized cost of collective action. If the revolution constraint given by equation (5.14) is not binding such that $\varepsilon \geq \varepsilon_2^{pe}$, the strong elite does not have a reason to democratize. Since there is no credible threat of revolution, they set their most preferred tax rate, which is zero. If the revolution constraint is binding for the strong elite such that $\varepsilon < \varepsilon_2^{pe}$, so that the revolutionary threat is credible, then the strong elite tries to prevent revolution by making redistribution or democratizing the regime.*

The strong elite can redistribute wealth to prevent revolution, only if $\varepsilon_2^{pe} > \varepsilon \geq \varepsilon_3^{pe}$, that is when (5.14) holds but inequality (5.15) does not hold. If inequality (5.15) holds, the strong elite cannot prevent revolution through redistribution. In this case, they can democratize. If inequality (5.16) holds, while inequality (5.15) does not hold, that is when $\varepsilon_3^{pe} > \varepsilon \geq \varepsilon_4^{pe}$, the strong elite can use democratization as a means to prevent revolution. If inequality (5.16) does not hold, that is when $\varepsilon \geq \varepsilon_4^{pe}$, even democratization cannot prevent revolution, and a revolution occurs.

From equation (5.12), it can be shown that the elite-poor inequality, κ_1, should be sufficiently low for an uprising by the poor alone to be feasible. That is to say, the poor should have sufficient resources. However, condition (5.14) puts an upward pressure on the elite-poor inequality in contrast to the condition describing the resource limitation of the poor. That is, κ_1 should be sufficiently high for the revolutionary threat to be credible. In the same way, a higher κ_1 also leads to higher critical levels for $\varepsilon_2^{pe}$and ε_3^{pe}, which shrinks the space to prevent the revolution by redistribution or democratization.

Intra-Elite Conflict Transitions

When $\varepsilon_1^{ie} > \varepsilon \geq \varepsilon_1^{pe}$, the poor cannot solve the collective action problem, and may form a coalition with the weak elite. If $\varepsilon \geq \varepsilon_1^{ie}$, solving the collective action problem of the poor is never feasible, and there is no revolutionary threat. Now assume that $\varepsilon_1^{ie} > \varepsilon \geq \varepsilon_1^{pe}$ holds. The poor or the weak elite will not be willing to undertake a revolution if the payoff from the revolution is lower than their income when the strong elite sets its ideal tax rate, which is zero. Therefore, one condition for the revolutionary threat from the coalition to be credible is that $V_{ie}^{p}(R,\mu) > y^p/(1-\beta)$. Otherwise, the strong elite anticipates that the poor will never revolt. This constraint can be written as

$$\Omega > \frac{\gamma^p - \theta(1-\mu) + \varepsilon\lambda^p(1-\mu)(1-\beta)}{\gamma^w(1-\mu)(1-\beta)},$$

which defines the relevant lower bound $\underline{\Omega}$. Another condition is $V_{ie}^{w}(R,\mu) > y^w/(1-\beta)$, that is, the weak elite should be able to solve the collective action problem of the poor. This leads to:

$$\Omega < \frac{(1-\theta)(1-\mu)\gamma^w}{\gamma^w(1-\mu)(1-\beta)},$$

which defines an upper bound $\overline{\Omega}$ for the transfer rate. The transfer to pose a revolutionary threat is feasible only if $\overline{\Omega} > \underline{\Omega}$, which leads to:

$$\varepsilon < \frac{1}{(1-\mu)\lambda^p(1-\beta)}[\gamma^s - \mu]. \tag{5.17}$$

We denote this critical level by ε_2^{ie}.

In a non-democracy, the best concession that the strong elite can offer is to promise the ideal tax rate to one of the member classes of the revolutionary coalition. If even these concessions cannot prevent at least one of the members from initiating the revolution, then the revolution is unpreventable. Note that the most preferred tax rate of the weak elite is zero, which leads to the revolution constraint where no redistribution takes place. When the strong elite makes a redistribution at a positive rate, the poor are better off, and they demand a larger post-revolutionary share from the strong elite's income which increases the lower bound. On the other hand, any positive rate of tax worsens the weak elite's situation. They demand less from the post-revolution allocation, which decreases the upper bound. Therefore, the best strategy that the strong elite can follow, to prevent revolution through redistribution, is to set the tax rate τ^p.[42] Then, the poor are willing to revolt, if $V^p_{ie}(R,\mu) > V^p(N,\varepsilon^H,\tau^p)$. By using the income share parameters, this constraint can be written as:

$$\begin{aligned}\Omega \;\; > \;\; & \frac{\gamma^p - \theta(1-\mu) + \varepsilon\lambda^p(1-\mu)(1-\beta)}{\gamma^w(1-\mu)(1-\beta)} \\ & + \frac{(1-\beta(1-q))(\tau^p(\lambda^w-\gamma^w) - C(\tau^p)\lambda^w)}{\gamma^w(1-\mu)(1-\beta)},\end{aligned}$$

defining a relevant lower bound $\underline{\Omega}$. Another condition is $V^w_{ie}(R,\mu) > V^w(N,\tau^p)$, that is the weak elite should be able solve the collective action problem of the poor. This leads to:

$$\Omega < \frac{(1-\theta)(1-\mu) - \gamma^w - (1-\beta(1-q))(\tau^p(\lambda^p-\gamma^p) - C(\tau^p)\lambda^p)}{\gamma^w(1-\mu)(1-\beta)}$$

which defines an upper bound $\overline{\Omega}$ for the transfer rate. The transfer can pose a revolutionary threat, only if $\overline{\Omega} > \underline{\Omega}$, which leads to:

$$\varepsilon < \frac{1}{(1-\mu)\lambda^p(1-\beta)}\left[\gamma^s - \mu - (1-\beta(1-q))B\right], \tag{5.18}$$

where $B = \tau^p(\gamma^s - \lambda^s) - C(\tau^p)(1-\lambda^s)$, which is the fraction of total income of the net transfer from the strong elite to the rest. We denote this critical level by ε^{ie}_3.

If a redistribution is not sufficient to prevent a revolution, the strong elite may democratize, which guarantees future redistributions. Applying the same method, we find the condition defining the critical level for the cost of collective action as:

$$\varepsilon < \frac{1}{(1-\mu)\lambda^p(1-\beta)}\left[\gamma^s - \mu - B\right]. \tag{5.19}$$

We denote this critical level by ε^{ie}_4.

[42] Note that we exclude a case in which the strong elite and the weak elite compromise on the basis of wealth transfers. The idea is that the strong elite obtains their wealth through rent seeking activities. This is the incentive for the weak elite to employ de facto power of the poor against the strong elite. As long as the weak elite leads a revolution, they will be better off by expropriating the strong elite.

We summarize the equilibrium for the case where the poor solve the collective action problem, only if the weak elite makes a transfer compensating the resource constraint of the poor:

Proposition 5.3. *Assume that the poor are not able to solve the collective action problem. If the revolution constraint given by equation (5.17) is not binding such that $\varepsilon \geq \varepsilon_2^{ie}$, the strong elite does not have a reason to democratize. Since there is no credible threat of revolution, they set their most preferred tax rate which is zero. If the revolution constraint is binding for the strong elite such that $\varepsilon < \varepsilon_2^{ie}$, the strong elite tries to prevent revolution by making a redistribution or democratizing the regime.*

The strong elite can redistribute to prevent revolution, only if $\varepsilon_2^{ie} > \varepsilon \geq \varepsilon_3^{ie}$, that is when (5.17) holds but inequality (5.18) does not hold. If inequality (5.18) holds, the strong elite cannot prevent revolution through a redistribution. In this case, they can democratize or repress. If inequality (5.19) holds, while inequality (5.18) does not hold, that is when $\varepsilon_3^{ie} > \varepsilon \geq \varepsilon_4^{ie}$, the strong elite can democratize in order to prevent revolution. If inequality (5.19) does not hold, that is when $\varepsilon \geq \varepsilon_4^{ie}$, even democratization cannot prevent revolution, and a revolution occurs.

Only if the elite-poor inequality is sufficiently high, so that the poor do not have sufficient resources to solve the collective action problem, they can form a coalition with the weak elite. In this case the weak elite finances the uprising by the poor, and have a share of the post-revolutionary income allocation. In contrast to the no coalition case, now the primary role in determining the equilibrium outcome belongs to the intra-elite inequality. The comparative statics with respect to the intra-elite inequality in the revolutionary coalition case are the same as the comparative statics with respect to the elite-poor inequality in the no coalition case. That is, a sufficiently low intra-elite inequality is required for the weak elite to be able to finance the uprising. However, it should also be high enough such that the weak elite is dissatisfied with the existing income allocation. In contrast to the elite-poor conflict case, the elite-poor inequality have two opposing effects on the critical levels. The additional effect is due to the reason that, in the revolutionary coalition case, the income of a poor agent is also important for the ability of the weak elite to be able to create an incentive for collective action. However, it can be shown that the net effect of κ_1 on the critical levels is in the same direction with its effect in the elite-poor conflict case.

The poor's trade off between acting alone and forming a coalition with the weak elite

When $\varepsilon < \varepsilon_1^{pe}$, the poor can revolt alone. Still, it is still possible that the poor may prefer to form a coalition with the weak elite. Now we will show that this is not the case. The net benefit of revolting alone instead of forming a coalition with the weak elite is

$$V_1^p(R,\mu) - V_2^p(R,\mu) = \frac{(1-\mu)\bar{y}(1-\theta-\Omega\gamma^w(1-\beta)}{\lambda^p(1-\beta)}.$$

This defines a critical level, $\hat{\Omega}$, such that when $\Omega < \hat{\Omega}$, the poor prefers to revolt alone. Remember that a revolutionary threat by the coalition is feasible only if $V_{ie}^w(R,\mu) > y^w/(1-\beta)$, that is the weak elite should be able solve the collective action problem of the poor. This leads to $\Omega < (1-\theta)(1-\mu)\gamma^w/(\gamma^w(1-\mu)(1-\beta))$, which defines an upper bound $\overline{\Omega}$ for the transfer rate. It can be shown that $\bar{\Omega} < \hat{\Omega}$. This means, in a parameter space where the poor is able to solve the collective action problem and the revolutionary threat by the coalition is credible, the poor always prefer to revolt alone. This leads to the following proposition:

Proposition 5.4. *When the poor are able to solve the collective action problem, the revolutionary threat, if it exists, is always posed by the poor alone. When the poor are not able to solve the collective action problem, the revolutionary threat, if it exists, can only be posed by the coalition between the poor and the weak elites.*

Proposition (5.4) also leads to some relational issues among our critical levels, which allows us to identify some interesting situations. Before explaining these situations, we will briefly discuss these properties. Firstly, note that $\varepsilon_2^{pe} > \varepsilon_3^{pe} > \varepsilon_4^{pe}$ always holds in the poor-elite conflict, consistent with the idea that the strong elite prefers no distribution at all, to preventing revolution via redistribution, and prefers these two strategies to democracy. However, the position of ε_1^{pe}, which describes the critical level for the collective action problem, with respect to other critical levels, is ambiguous. For example, we may have $\varepsilon_1^{pe} < \varepsilon_2^{pe}$. This means, whenever the poor is able to solve the collective action problem, there must exist a revolutionary threat by the poor alone. The same situation arises in the intra-elite conflict case. While $\varepsilon_2^{ie} > \varepsilon_3^{ie} > \varepsilon_4^{ie}$ always holds, the position of ε_1^{ie} is ambiguous. Therefore, our parameter space may lead to a situation where some cases described by proposition (5.1) and (5.2) never appear.

Secondly, $\varepsilon_2^{pe} > \varepsilon_2^{ie}$ always holds. That is, if the revolution constraint is binding for the coalition, then it must also be binding for an uprising by the poor alone. In such a case, the only thing that may drive the poor to form a coalition is their inability to solve the collective action problem. Similarly, $\varepsilon_3^{pe} > \varepsilon_3^{ie}$ and $\varepsilon_4^{pe} > \varepsilon_4^{ie}$ always hold. The first inequality indicates that if the revolution is prevented by redistribution in the poor-elite conflict, then a revolutionary threat by the coalition must always be met by a redistribution. The second inequality implies that if democratization prevents revolution in the poor-elite conflict, then democratization must prevent revolution in the intra-elite conflict case. In such a setting, we are able to identify the conditions when intra-elite conflict is never a threat to non-democracy.

5.3. Elite Unification, Income Inequality, and Consolidation of Democracy

Acemoglu and Robinson (2006) explain the reason for the consolidation of democracy in Britain with relatively low levels of inequality between the unified elite class and the poor capable of acting collectively. When inequality is low, following a democratic transition, the poor demand less redistribution. Thus, the burden of democracy on the elites is lower, making coups less attractive for the elites. On the other hand, Burton and Higley (1987) argue that the elite unification occurred between the Tories and the Whigs in Britain, around the 1840s, in response to civil wars "unleashing the leveling social revolutionary tendencies". This is considered as the main reason for the consolidation of British democracy. However, they do not relate the conflicting interests of elites with the occurrence of social movements. By proposing the inability of elite factions to mobilize the poor, as a precondition for elite unification, our model relates elite unification to democratic transitions through income distribution.

In our model, an intra-elite conflict arises from the presence of discontented elite factions which are able to mobilize the masses and create a revolutionary threat. However, the threat by the weak elite is never credible, if our parameter space is such that ε_1^{pe} is higher than ε_2^{ie}, which means that whenever intra-elite conflict leads to a revolutionary threat, the poor are able to act collectively, and they do so by proposition (5.4). In such a society, intra-elite conflict never threatens the non-democratic regime. In terms of inequality parameters, the condition $\varepsilon_1^{pe} > \varepsilon_2^{ie}$ leads to:

$$\mu > \frac{\kappa_1\kappa_2 - (1-\kappa_1)(1-\beta)}{1-(1-\kappa_1)(1-\beta)}. \tag{5.20}$$

We denote the critical level of μ defined by condition (5.20) by μ^*. A simple comparative static analysis reveals the following result:

Proposition 5.5. *$\mu^*$is increasing in both κ_1 and κ_2:*

$$\frac{\partial \mu^*}{\partial \kappa_1} = \frac{\kappa_2 + (1-\beta)(1-\kappa_2)}{[1-(1-\kappa_1)(1-\beta)]^2} > 0, \quad \frac{\partial \mu^*}{\partial \kappa_2} = \frac{\kappa_1}{[1-(1-\kappa_1)(1-\beta)]^2} > 0.$$

Therefore, condition (5.20) is more likely to hold, when both intra-elite and poor-elite inequality is lower. That is, a society with a unified elite structure must have a more equal income distribution compared to a society where a threat of intra-elite conflict does exist.

In our model, a unified elite structure, rather than being a result of strategic movements of contending elites, arises from the elimination of the ability of discontented elite factions to mobilize the masses. Proposition (5.5) indicates that this is possible if the income distribution is sufficiently equal. Therefore, the proposed preconditions for a democratic transition to create a consolidated one are consistent: A unified elite structure will only exist in sufficiently equal societies.

5.4. Revolutions

In this section, our main goal is to analyze equilibrium revolutions, arising at where democratization is not sufficient to meet revolutionary threat. Specifically, we make a comparative analysis of the events of 1848 and 1875 in France, 1917 in Russia, and 1911-1949 in China. While doing this, we restrict our attention to intra-elite conflict case, since all of these revolutions replaced the monarchial elites of old-regimes with emerging elites.[43] Therefore, throughout this section, we assume that $\varepsilon_1^{ie} > \varepsilon \geq \varepsilon_1^{pe}$ holds. That is, the cost of collective action prevents the poor to collectively act, but

[43] We ignore the emprically irrelevant case where the poor-elite conflict leads to revelution.

allows a coalition with the weak elite in order to pose a revolutionary threat. The post-revolution outcomes in these societies are not the same. Revolution in France resulted in some political reforms towards democracy, whereas, in China and Russia, revolutions created new non-democracies. Following, we will analyze these distinct revolutionary paths.

Why do some revolutionary mass movements result in consolidated non-democracies, such as in Russia and China in the first half of the twentieth century? In these cases, the revolutions created new non-democracies with new elites even stronger than the elites of the previous regimes. Based on this observation, we assume that these revolutions are undertaken by the coalition of discontent elite factions and the masses. Furthermore, we assume the revolutionary threat cannot be met by the strong elite. That is, $\varepsilon_4^{ie} > \varepsilon$ holds. Now, assume that when a revolution takes place, we are in a non-democracy where the weak elite has all the political power. Moore (1966) argues that totalitarian regimes arise from the domination of an elite class. However, do the weak elite has the chance to maintain the post-revolutionary non-democracy? Our answer is yes. In intra-elite conflict revolutions, if the weak elite can offer to the poor a combination of current transfer to solve the collective action problem (Ω), and a post-revolution income share, θ, such that the offer is sufficient to lead the poor to a revolution, but the post-revolution income share of the poor is not sufficient to act collectively in a post-revolution non-democracy, then the post-revolutionary regime is a non-democracy where all political power is concentrated in the hands of the weak elite. To derive the condition where such an outcome is possible, first note that, if the weak elite is able to offer a post-revolution income share (θ), which is smaller than the actual income share of the poor (γ^p), so that the poor are not able to solve the collective action problem in the post-revolution period, then existence of such an equilibrium is guarenteed. Now assume that the weak elite makes the highest possible transfer, namely $\Omega = 1$. If, in this case, the weak elite can offer a $\theta < \gamma^p$, just making the poor indifferent between revolution and democracy, then consolidation of post-revolution non-democracy can always be achieved. Now, we can state the proposition.

Proposition 5.6. *If the following condition holds*

$$\frac{B^p}{(1-\beta)} < (\gamma^w - \varepsilon\lambda^p)(1-\mu)^2, \tag{5.21}$$

then the weak elite always offers a combination of Ω and θ, such that the poor revolt, but are not able to solve the collective action problem in the post-revolution non-democracy.

Proof. See Appendix *Q.E.D.*

In inequality (5.21), $B^p = \tau^p(\lambda^p - \gamma^p) - C(\tau^p)\lambda^p$ is the net per period redistribution that the poor would get if democratization takes place. On the other hand, the expression on the right hand side measures the ability of the weak elite to mobilize the masses, or equivalently, it is what the poor require in order to take action. Both sides of inequality (5.21) is increasing in elite-poor inequality, κ_1, and the net effect is ambigious. However, higher intra-elite inequality, κ_2, decreases the likeliness of inequality (5.21) to hold, since it means that the weak elite is economically less powerful to make direct transfers. In such a case, in order to mobilize the masses, they have to offer a higher post-revolutionary share, which makes the poor more likely to be able to collectively act in the post-revolutionary period. Hence, establishment of non-democracy is more difficult when κ_2 is higher.

5.5. Historical Evidence

5.5.1. Democratic Transitions in Britain and Denmark

In the first half of the 19^{th} century, a series of political reforms took place in both Britain and Denmark. In Britain, the franchise was first extended with the Reform Act of 1832, and gradually expanded to 64% of adult males by the Reform Acts of 1867 and 1884. In Denmark, universal male suffrage was introduced in 1949. In line with our theoretical model, we analyze the form of the conflict leading to these two cases in two dimensions: The elite structure and the collective action problem of the masses.

Elite unification as a result of the Glorious Revolution (Burton et al., 1992) placed Britain on a rapid industrialization path. This prepared the conditions for the emergence of a collectively active working class at the end of the 18^{th} century. For example, from 1800 to 1850 the population of the ten largest cities in Britain are doubled. The social unrest in this era showed itself in the Luddite Riots, the Spa Fields Riots, the Peterloo Massacre, and the Swing Riots. The Chartist movement of the working class between

1838 and 1848 made the necessity of further political reform obvious for the elites. Thus the political reforms in Britain are an example of poor-elite conflict transitions, where the collectively active poor is a threat to the unified elites.

The introduction of universal male suffrage in Denmark in 1849 can be explained by a series of events mainly initiated by discontented elites. Industrialization started in Denmark after 1850. Hence, as opposed to the British case, it is hard to identify a collectively active industrial working class until that time. The upheavals were partly due to the bourgeoisie-led peasant movement (Collier, 1999), in addition to the strong nationalist movements in the Schleswig and Holstein Duchies. In both cases conflict arose within the elite, and the masses was mobilized by discontented elites. These upheavals resulted in extension of franchise which was not actually the intention of the discontented elites. Indeed, soon thereafter Schleswig and Holstein joined the German Confederation. The case of Denmark illustrates intra-elite conflict transitions where the masses are mobilized by the discontented elites.

5.5.2. Democratic Consolidation

Following, we analyze the implications of proposition (5.5) on the political reforms in the 19^{th} century Britain. However, initially, it is important to highlight the historical process preparing the conditions which leads to these reforms. Compared to the long history of autocratic states, democratic transitions, emerging in the 19^{th} century, are very recent events. So, what are the underlying reasons behind this fact? Why, for example, was the British democracy not established before the 19th century?

Our model places the ability of the masses to act collectively at the center of the analysis, which can change in the pre-transition period independent of the income distribution. Field and Higley (1980) highlight the passive role of the non-elites in the seventeenth and eighteenth century politics of Britain, and how this situation changed in the nineteenth century as the working class evolved from peasantry to an industrial working class. At this point, we underline the implications of this situation for the collective action problem. While the peasantry class has mostly private property orientations, the industrial working class has the potential to realize their common interests as in nineteenth-century Britain. Specifically, according to our model, what have changed in the pre-transition period is a decline in the cost of collective action. Such a decline can be a result of structural changes associated with industrialization such as higher urbanization, more specialized labor requirements, and the poor in a more impersonalized production process, which all leads the poor to realize the conflict over the income redistribution, and ease collective action.

According to our model, another element of change in the period preceding 19^{th} century Britain was a structural change within the elite class. Before the industrial revolution, any conflict within the elite class was mostly due to the rents over land. The equilibrium between elites was a result of balanced military forces. However, following the decades of the discovery of the new world and the industrial revolution, there had been a crucial change within the elite class. Most importantly, saving and investing were not a privilege of a dense elite class. Capital ownership had broadened to a larger group. These new capital owners, the so-called bourgeoisie, had different interests to those of the existing elites. In order to protect their savings and investments, they needed different political institutions to the existing traditional institutions. Lachmann (1990) gives priority to the intra-elite conflict in British transition from feudalism to capitalism experienced from the 16^{th} to the 18^{th} century. He argues that, although the interests of the elites are determined by inter-class relations, their capacity to pursue their interests is primarily restricted by intra-elite relations. According to Lachmann (1990), only when a unified elite rules, can the situation be analyzed as a struggle between the ruling and the producing classes. As Lachmann (1990), our model describes 19^{th} century Britain with the case where the main determinant of the regime is the elite-poor conflict. Burton et al. (1992) argues that elite unification in Britain was a result of the Glorious Revolution. In proposition (5.5), we show that the conditions leading to a unified elite structure in Britain are those which are argued to consolidate democracy.

In our model, we analyze the role of both the collective action problems and the elite structure on democratic transitions, although we do not explain how they evolve over time. Moreover, we argue that these two possible explanations are interrelated. Democratic transition in Britain can be explained by such an analysis. Our model puts forward two characteristics of the pre-transition period. Firstly, elite unification was achieved by elimination of potential mass mobilization by the discontented elite factions. Secondly, rapid industrialization reduced the costs of collective action, possibly through higher urbanization, more education, and impersonalized production processes, paving the way for the poor to act collectively with greater ease. According to proposition (5), such an elite-poor conflict transition, where elite unification has been achieved and the poor are able to solve the collective action problem, is more likely in relatively equal societies. Therefore, the two explanations for the consolidation of democracy in Britain, a low income inequality and a unified elite structure, are consistent preconditions.

In contrast, when condition (5.20) does not hold, intra-elite conflict transitions are possible. Proposition (5.5) indicates that higher inequality makes this situation more likely. Indeed, while in Britain the income share of the richest 14% was 0.44 in 1688 (Maddison, 2007), in Denmark the top decile's share of income was 0.50 in 1870 (Morrisson, 1999). This is consistent with our view that the 1849 reform in Denmark is an example of intra-elite conflict transition. Moreover, these reforms in Denmark were not long-lasting, in line with the arguments of Acemoglu and Robinson (2006).

5.5.3. Revolutionary Transition to Democracy

In the previous sections, we argued that the 1849 Reform in Denmark took place in response to an intra-elite conflict. Why did Denmark succeed meeting the revolutionary threat by political reforms, while monarchial elites were overthrown in France, Russia, and China? According to our model, meeting the revolutionary threat by democratization is not possible, if $\varepsilon_4^{ie} > \varepsilon$ holds. Simple comparative statics on ε_4^{ie} reveals that higher intra-elite inequality and elite-poor inequality make revolutions more likely. Indeed, according to Morrisson (1999), the top decile income share in France from 1788 to 1864 was ranging between 0.50 and 0.55, indicating a very high level of inequality.

When the condition in proposition (5.6) does not hold, the weak elite may not be able to seize power without satisfying the demands of the masses, if a non-democracy were established following the revolution.[44] In this case, the weak elite may establish democracy following the revolution. Such a case is well exemplified by the reforms in 19^{th} century France. Following the French Revolution, elites were divided along the monarchist and the republican camp. The democratic reforms of 1848 was a result of the republican challenge to the monarchists. The republican movement was a bourgeoisie-led movement in a coalition with the working class (Ghosal and Proto, 2009). However, the republicans

[44]Note that, in this case, the post-revolutionary game is exactly same with case of elite-poor conflict. For this reason, we do not go into analytical details.

were socially conservative (Collier, 1999) in the sense that they were against the universal male suffrage. However, as they seized power, universal male suffrage was introduced, not because they favored it, but because the working class was a class-conscious actor in the pre-reform movements. Indeed, in the July revolution of 1830, the working class was an active participant. While the peasantry has a private property orientation which can be met by land reforms, an industrial working class has the potential to realize their common interests in future redistributions. That is, as opposed to the Russian and Chinese cases, the elite-led mass movement was motivated by the future gains (which is represented by the left hand side of inequality (5.21)). The new elites seizing power were constrained by pressure from below following the revolution. As Aminzade (1993) argues, working class were the revolutionary force that put the republicans in power in 1848, and working class pressure forced republicans accept universal male suffrage.

The same nature of inter and intra class conflicts are valid for the reforms following 1870. Collier (1999) argues that "the democratic constitution of 1875 and the resolution of the deadlock in 1877 can be understood in terms of both divisions within the governing monarchial elites and rapid political recovery of the republican movement". The political reform was again in response to working class pressure, particularly the Paris Commune (Acemoglu and Robinson, 2006).

5.5.4. Post-revolutionary Non-democracy

In the pre-revolutionary period, both Russian and Chinese societies were predominantly agrarian, and peasant revolts played a crucial role in these revolutions (Skocpol, 1979).[45] According to Moore (1966), peasants were the "dynamite" bringing the old-regimes down. However, neither Skocpol nor Moore do not focus on the obvious answer to the question that why the peasant revolts did not appear in the post-revolution period. As Skocpol (1979) argues, main motivation of peasant movements is land allocation. While land reforms were contradictory with the monarchial character of the old-regimes[46], it was a possible concession for the post-revolutionary regimes, since the main concern of

[45]According to Skocpol (1979), Russian working-class movement played a role only in shaping the outcome of the revolution, however, major factor breaking the old-regime down was peasant movements.

[46]Skocpol (1979) states "Together the extensiveness and anti-landlord focus of the revolutionary peasant revolts created decisive constraints at the societal level on the range of sociopolitical options available to elites contending for national power."

revolutionary leaders were creating industrialized states. Indeed, both Russian and Chinese post-revolutionary regimes engaged in highly extensive land reforms, and took a industrialized development path. This is what we refer as the pre-revolution agreements on the post-revolution income allocation in our model. Land reforms were reachable outcomes for the peasantry. This is the opportunity arised with industrial revolution, which opened the way for the revolutionary leaders to make a credible promise about the post-revolution income allocation via land reforms. This also explains why the previous peasant uprisings never led to revolutions. In other words, it was possible for the weak elites to mobilize the masses by keeping their post-revolution income share at a minimum level, which, in turn, prevents collective action following the revolution. Therefore, in our point of view, preponderence of peasantry's role in these revolutions is the reason which made a non-democratic regime possible in the post-revolution era.

5.6. Conclusion

In this paper, the intra-elite conflict is explained as an inherent character of any non-democratic society. We show that it does not appear as a threat in societies with relatively equal income distribution. Hence, in these societies the only determinant of a regime transition is the conflict between the masses and a unified elite. This result depends on the idea that when the masses is resource-constrained in solving the collective action problem, an elite faction which is discontented with the existing regime, can employ de facto power of the poor. Hence, intra-elite conflict and mass movements are considered as related issues effecting regime transitions.

Furthermore, we incorporate our intra-elite conflict model to explain the differential outcomes of revolutions. We show that the outcome of revolutions depends on the combination of current transfers to solve the collective action problem of the masses, and the post-revolution income allocation. When intra-elite inequality is high, this combination either cannot mobilize the mass or cannot keep the poor in post-revolutionary era collectively inactive. In this case, emerging elites have to make concessions.

As a future research, the role of collective action in our model can be incorporated into models of regime transitions which aim to explain the expected decline in the cost of collective action with the industrialization in the pre-transition periods. Such a model can construct the link between growth via industrialization and political transitions.

5.A. Appendix

5.A.1. Proof of proposition (5.6)

In order to provide a better intuition, we start with analyzing proposition 5.3. All cases in intra-elite conflict, summarized in proposition 5.3, are outcomes of the possibility of a cooperation between the weak elite and the poor, plus the ability of the strong elite to prevent revolution. We will analyze the most strict case, which is about whether democracy can prevent a revolution. Results can be generalized to other statements of proposition 5.3.

The weak elite prefers to initiate a revolution, if $V^w_{ie}(R, \mu) > V^w(D)$, which gives

$$\Omega < \frac{(1-\theta)(1-\mu) - \gamma^w - B(w)}{\gamma^w(1-\mu)(1-\beta)},$$

where $B(w) = \tau^p(\lambda^p - \gamma^p) - C(\tau^p)\lambda^p$. Secondly, the poor prefers revolution to democracy, if $V^p_{ie}(R, \mu) > V^p(D)$, which gives

$$\Omega > \frac{\gamma^p - \theta(1-\mu) + \varepsilon\lambda^p(1-\mu)(1-\beta) + B(p)}{\gamma^w(1-\mu)(1-\beta)},$$

where $B(p) = (\tau^p(\lambda^w - \gamma^w) - C(\tau^p)\lambda^w)$. These conditions, as well as the conditions for other cases in proposition 5.3, can be summarized as

$$\Omega \;<\; -a\theta + b, \tag{5.22}$$

$$\Omega \;>\; -a\theta + c, \tag{5.23}$$

where, in the case of democracy-revolution trade-off,

$$\begin{aligned} a &= \frac{1-\mu}{\gamma^w(1-\mu)(1-\beta)}, \\ b &= \frac{1-\mu-\gamma^w - B(w)}{\gamma^w(1-\mu)(1-\beta)}, \\ c &= \frac{\gamma^p + \varepsilon\lambda^p(1-\mu)(1-\beta) + B(p)}{\gamma^w(1-\mu)(1-\beta)}. \end{aligned}$$

Inequality (5.22) is the resource constraint of the weak elite which defines the upper bound, $\bar{\Omega}$. Inequality (5.23) is the condition for the poor to choose revolution instead of some concession by the strong elite, which defines a lower bound, $\underline{\Omega}$. It can be verified that all coefficients of this inequality system are positive. The condition in the text, $\bar{\Omega} > \underline{\Omega}$, for a revolution takes place is, therefore, $b > c$. Now we go on with the proof of proposition 5.6 by assuming $b > c$.

$$\theta < \varepsilon\lambda^p. \tag{5.24}$$

Note, we have to impose $\varepsilon \geq \gamma^p/\lambda^p$, since intra-elite conflict arises only when the poor is unable solve the collective action problem. In terms of preventing post-revolutionary collective action, the most strict situation for the weak elite arises when this condition holds with equality, $\varepsilon = \gamma^p/\lambda^p$. Substituting in (5.24) gives $\theta < \gamma^p$.

Now assume that the weak elite makes no transfer to the poor, $\Omega = 0$. Substituting in inequality (5.22), gives $\theta > c/a$, which contradicts with the condition that $\theta < \gamma^p$, since these conditions hold simultaneously only when $\gamma^p > c/a$, which never holds. Therefore, an Ω which satisfies all the required conditions must be higher than zero. Secondly, assume that the weak elite makes the highest possible transfer, $\Omega = 1$. Substituting in inequality (5.22), gives $\theta > (c-1)/a$. In order the condition, $\theta < \gamma^p$, also to hold, we must have $\gamma^p > (c-1)/a$. This is the condition in proposition (5.6). This suggest that there exists a $\theta < \gamma^p$ when $\Omega = 1$, which is sufficient to activate the poor for a revolution, and prevents post-revolution collective action. Only remaining issue is to check whether this combination is beneficial for the weak elite. Substituting $\Omega = 1$ in the resource constraint of the weak elite, inequality (5.22) gives $\theta < b/a$, which always holds, since $\theta > (c-1)/a$ and $b > c$ by assumption. Therefore, there exists a $\Omega > \Omega^* = -a\theta + c$ for which there exists a $\theta < \gamma^p$. This completes the proof.

Bibliography

Acemoglu, D. and J. Robinson (2000). Why did the west extend the franchise? democracy, inequality, and growth in historical perspective. *The Quarterly Journal of Economics 115*(4), 1167–1199.

Acemoglu, D. and J. Robinson (2001). A theory of political transitions. *American Economic Review*, 938–963.

Acemoglu, D. and J. Robinson (2006). *Economic origins of dictatorship and democracy.* Cambridge Univ Pr.

Aghion, P., A. Dechezlepretre, D. Hemous, R. Martin, and J. Van Reenen (2010). Carbon taxes, path dependency and directed technical change: Evidence from the auto industry. Technical report, Working Paper.

Aghion, P., M. Dewatripont, and P. Rey (1997). Corporate governance, competition policy and industrial policy. *European Economic Review 41*(3), 797–805.

Aghion, P. and R. Griffith (2005). *Competition and growth: reconciling theory and evidence.* The MIT Press.

Ambec, S. and P. Barla (2002). A theoretical foundation of the porter hypothesis. *Economics Letters 75*(3), 355–360.

Aminzade, R. (1993). *Ballots and barricades: class formation and republican politics in France, 1830-1871.* Princeton Univ Pr.

Anderson, R., J. Hair, R. Tatham, and W. Black (2006). *Multivariate data analysis.* Pearson.

André, F., P. González, and N. Porteiro (2009). Strategic quality competition and the porter hypothesis. *Journal of Environmental Economics and Management 57*(2), 182–194.

Auffhammer, M. and R. T. Carson (2008). Forecasting the path of china's co2 emissions using province-level information. *Journal of Environmental Economics and Management 55*(3), 229–247.

Auffhammer, M. and R. Steinhauser (2012). Forecasting the path of us co2 emissions using state-level information. Review of Economics and Statistics *94*(1), 172–185.

Azomahou, T., F. Laisney, and P. N. Van (2006). Economic development and co2 emissions: A nonparametric panel approach. *Journal of Public Economics 90*, 1347–1363.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. Econometrica *70*(1), 191–221.

Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. Econometrica *72*(4), 1127–1177.

Banerjee, A., M. Marcellino, and C. Osbat (2004). Some cautions on the use of panel methods for integrated series of macroeconomic data. The Econometrics Journal *7*(2), 322–340.

Berle, A. and G. Means (1932). *The modern corporation and private property.* Transaction Pub.

Boden, T.A., G. M. and R. Andres. (2013). Global, regional, and national fossil-fuel co2 emissions. Technical report, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Boden, T. A., G. Marland, and R. Andres (1995). Estimates of global, regional, and national annual carbon dioxide emissions from fossil-fuel burning, hydraulic cement production, and gas flaring: 1950–1992. Technical report, Oak Ridge National Lab., TN (United States).

Bolt, J. and J. L. van Zanden (2013). The first update of the maddison project; reestimating growth before 1820. Technical report, Maddison Project Working Paper 4.

Breitung, J. (2000). The local power of some unit root tests for panel data. In *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels, JAI.*

Brunnermeier, S. and M. Cohen (2003). Determinants of environmental innovation in us manufacturing industries. *Journal of environmental economics and management 45*(2), 278–293.

Burton, M., R. Gunther, and J. Higley (1992). Introduction: elite transformations and democratic regimes. *Elites and democratic consolidation in Latin America and Southern Europe 1*, 37.

Burton, M. and J. Higley (1987). Elite settlements. *American Sociological Review*, 295–307.

Chang, Y., J. Park, and P. Phillips (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. Econometrics Journal *4*, 1–36.

Chatfield, C. (2002). *Time-series forecasting.* Chapman and Hall/CRC.

Chertow, M. (2000). The IPAT Equation and Its Variants. Journal of Industrial Technology *4*, 13–29.

Cho, M. (1998). Ownership structure, investment, and the corporate value: an empirical analysis. *Journal of Financial Economics 47*(1), 103–121.

Choi, I. (2001). Unit root tests for panel data. Journal of International Money and Finance *20*(2), 249–272.

Choi, I. and E. Kurozumi (2012). Model selection criteria for the leads-and-lags cointegrating regression. Journal of Econometrics.

Clarke, L., J. Edmonds, H. Jacoby, H. Pitcher, J. Reilly, and R. Richels (2007). Scenarios of greenhouse gas emissions and atmospheric concentrations. *US Department of Energy Publications*, 6.

Collier, R. (1999). *Paths toward democracy: The working class and elites in Western Europe and South America.* Cambridge Univ Pr.

Commoner, B. (1972). The environmental cost of economic growth. *Chemistry in Britain 8*(2), 52.

De Bruyn, S., J. van den Bergh, and J. Opschoor (1998). Economic growth and emissions: reconsidering the empirical basis of environmental kuznets curves. Ecological Economics *25*(2), 161–175.

De Hoyos, R. and V. Sarafidis (2007). Testing for cross-sectional dependence in panel-data models. Stata Journal *6*(4), 482–496.

Demsetz, H. and K. Lehn (1985). The structure of corporate ownership: Causes and consequences. *The Journal of Political Economy 93*(6), 1155–1177.

Demsetz, H. and B. Villalonga (2001). Ownership structure and corporate performance. *Journal of Corporate Finance 7*(3), 209–233.

Dietz, T. and E. A. Rosa (1994). Rethinking the environmental impacts of population, affluence and technology. *Human Ecology Review 1*, 277–300.

Dijkgraaf, E. and H. Vollebergh (2005). A test for parameter homogeneity in co sub (2) panel ekc estimations. Environmental & Resource Economics *32*(2), 229–239.

Ehrlich, P. R., J. P. Holdren, et al. (1971). Impact of population growth. *Science 171*(3977), 1212–1217.

Field, G. and J. Higley (1980). *Elitism.* Routledge & Kegan Paul London.

Fisher, R., S. Genetiker, R. Fisher, S. Genetician, G. Britain, R. Fisher, and S. Généticien (1970). *Statistical methods for research workers*, Volume 14. Oliver and Boyd Edinburgh.

Fujino, J., R. Nair, M. Kainuma, T. Masui, and Y. Matsuoka (2006). Multi-gas mitigation analysis on stabilization scenarios using aim global model. *Energy Journal 27.*

Gabel, H. and B. Sinclair-Desgagné (1997). *The firm, its routines, and the environment.*

Galeotti, M., M. Manera, and A. Lanza (2009). On the robustness of robustness checks of the environmental kuznets curve hypothesis. Environmental and Resource Economics *42*(4), 551–574.

Gengenbach, C., F. Palm, and J. Urbain (2006). Cointegration testing in panels with common factors*. Oxford Bulletin of Economics and Statistics *68*(s1), 683–719.

Ghosal, S. and E. Proto (2009). Democracy, collective action and intra-elite conflict. *Journal of Public Economics 93*(9), 1078–1089.

Giacomini, R. and C. W. Granger (2004). Aggregation of space-time processes. *Journal of econometrics 118*(1), 7–26.

Granger, C. W. and P. Newbold (1976). Forecasting transformed series. Journal of the Royal Statistical Society. Series B (Methodological), 189–203.

Greaker, M. (2003). Strategic environmental policy; eco-dumping or a green strategy? *Journal of Environmental Economics and Management 45*(3), 692–707.

Grossman, G. and A. Krueger (1991). Environmental impacts of a north american free trade agreement. Technical report, National Bureau of Economic Research.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. The Econometrics Journal *3*(2), 148–161.

Hamamoto, M. (2006). Environmental regulation and the productivity of japanese manufacturing industries. *Resource and energy economics 28*(4), 299–312.

Hart, O. (1983). The market mechanism as an incentive scheme. *The Bell Journal of Economics*, 366–382.

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics 115*(1), 45–97.

Hermalin, B. and M. Weisbach (1988). The determinants of board composition. *The RAND Journal of Economics*, 589–606.

Higley, J. and M. Burton (1989). The elite variable in democratic transitions and breakdowns. *American Sociological Review*, 17–32.

Hijioka, Y., Y. M. H. N. M. M. and M. Kainuma (2008). Global ghg emissions scenarios under ghg concentration stabilization targets. *Journal of Global Environmental Engineering 13*, 97–108.

Himmelberg, C., R. Hubbard, and D. Palia (1999). Understanding the determinants of managerial ownership and the link between ownership and performance. *Journal of financial economics 53*(3), 353–384.

Holderness, C., R. Kroszner, and D. Sheehan (1998). Were the good old days that good? changes in managerial stock ownership since the great depression. Technical report, National Bureau of Economic Research.

Holtz-Eakin, D. and T. Selden (1995). Stoking the fires? co2 emissions and economic growth. Journal of Public Economics *57*(1), 85–101.

Horvath, R. (1997). Energy consumption and the environmental kuznets curve debate. *Department of Geography, University of Sydney, Sydney NSW*.

Im, K., M. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. Journal of Econometrics *115*, 53–74.

IPCC (2000). Special report on emissions scenarios (sres). a special report of working group iii of the intergovernmental panel on climate change.

IPCC (2008). Climate change 2007. synthesis report. contribution of working groups i, ii and iii to the fourth assessment report.

Jaffe, A., R. Newell, and R. Stavins (2003). Technological change and the environment. *Handbook of environmental economics 1*, 461–516.

Jaffe, A. and K. Palmer (1997). Environmental regulation and innovation: a panel data study. *Review of Economics and Statistics 79*(4), 610–619.

Jensen, M. and W. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics 3*(4), 305–360.

Johnstone, N., I. Hascic, and D. Popp (2010). Renewable energy policies and technological innovation: Evidence based on patent counts. Environmental and Resource Economics *45*(1), 133–155.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. Journal of Econometrics *90*(1), 1–44.

Kao, C. and M. Chiang (2001). On the estimation and inference of a cointegrated regression in panel data.

Kapetanios, G., M. Pesaran, and T. Yamagata (2011). Panels with non-stationary multifactor error structures. Journal of Econometrics *160*(2), 326–348.

Karlsen, H., T. Myklebust, and D. Tjøstheim (2007). Nonparametric estimation in a nonlinear cointegration type model. Annals of Statistics *35*(1), 252–299.

Kejriwal, M. and P. Perron (2008). Data dependent rules for selection of the number of leads and lags in the dynamic ols cointegrating regression. Econometric Theory *24*(5), 1425.

Komen, M., S. Gerking, and H. Folmer (1997). Income and environmental r&d: empirical evidence from oecd countries. Environment and Development Economics *2*(04), 505–515.

Kuznets, S. (1955). Economic growth and income inequality. *The American economic review*, 1–28.

Lachmann, R. (1990). Class formation without class struggle: An elite conflict theory of the transition to capitalism. *American Sociological Review*, 398–414.

Lanjouw, J. and A. Mody (1996). Innovation and the international diffusion of environmentally responsive technology. *Research Policy 25*(4), 549–571.

Levin, A., C. Lin, and C. Chu (2002). Unit root tests in panel data: asymptotic and finite-sample properties. Journal of Econometrics *108*, 1–24.

Loderer, C. and K. Martin (1997). Executive stock ownership and performance tracking faint traces. *Journal of Financial Economics 45*(2), 223–255.

Lükepohl, H. (1987). *Forecasting aggregated vector ARMA processes*, Volume 284. Springer.

Lütkepohl, H. (1984). Linear transformations of vector arma processes. Journal of Econometrics *26*(3), 283–293.

Lütkepohl, H. (2006). Forecasting with varma models. Handbook of Economic Forecasting *1*, 287–325.

Lütkepohl, H. and F. Xu (2012). The role of the log transformation in forecasting economic variables. Empirical Economics *42*(3), 619–638.

Maddala, G. and S. Wu (1999). A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and statistics *61*(S1), 631–652.

Maddison, A. (2007). *Contours of the world economy, 1-2030 AD: essays in macro-economic history.* Oxford University Press, USA.

Maddison, A. (2009). Statistics on world population, gdp and per capita gdp, 1-2006 ad, 2008.

Mammen, E., O. Linton, and J. Nielsen (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Annals of Statistics *27*(5), 1443–1490.

Mammen, E. and B. Park (2005). Bandwidth selection for smooth backfitting in additive models. The Annals of Statistics *33*(3), 1260–1294.

Marcellino, M., J. H. Stock, and M. W. Watson (2003). Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review 47*(1), 1–18.

Marland, G., T. Boden, and B. Andres (2009).

Martinez-Zarzoso, I. and A. Bengochea-Morancho (2004). Pooled mean group estimation of an environmental kuznets curve for co< sub> 2</sub>. Economics Letters *82*(1), 121–126.

McConnell, J. and H. Servaes (1990). Additional evidence on equity ownership and corporate value. *Journal of financial Economics 27*(2), 595–612.

Melenberg, B., H. Vollebergh, and E. Dijkgraaf (2011). Grazing the commons: Global carbon emissions forever?

Melenberg, B., H. Vollebergh, and S. Sen (2014). Pairwise differencing forecast of global carbon dioxide emissions: China vs. technological effects. Technical report, Tilburg University.

Millimet, D., J. List, and T. Stengos (2003). The environmental kuznets curve: Real progress or misspecified models? Review of Economics and Statistics *85*(4), 1038–1047.

Mohr, R. (2002). Technical change, external economies, and the porter hypothesis. *Journal of Environmental economics and management 43*(1), 158–168.

Mohr, R. and S. Saha (2008). Distribution of environmental costs and benefits, additional distortions, and the porter hypothesis. *Land Economics 84*(4), 689–700.

Moon, H. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. Journal of Econometrics *122*(1), 81–126.

Moon, H., B. Perron, and P. Phillips (2007). Incidental trends and the power of panel unit root tests. Journal of Econometrics *141*(2), 416–459.

Moore, B. (1966). Social origins of dictatorship and democracy. *Boston, 1966. Morgan.*

Morck, R., A. Shleifer, and R. Vishny (1988). Management ownership and market valuation:: An empirical analysis. *Journal of financial economics 20*, 293–315.

Morrisson, C. (1999). Historical evolution of income distribution in western europe. *Handbook of Income Distribution, Amsterdam: North-Holland.*

Mosca, G., H. Kahn, and A. Livingston (1939). *The ruling class:(Elementi di scienza politica).* McGraw-Hill Book Co. Inc.

Moss, R. H., J. A. Edmonds, K. A. Hibbard, M. R. Manning, S. K. Rose, D. P. Van Vuuren, T. R. Carter, S. Emori, M. Kainuma, T. Kram, et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature 463*(7282), 747–756.

Muller-Furstenberger, G. and M. Wagner (2007). Exploring the environmental kuznets hypothesis: Theoretical and econometric problems. Ecological Economics *62*(3), 648–660.

Nelson, R. and S. Winter (1982). *An evolutionary theory of economic change.* Belknap press.

Nielsen, J. and S. Sperlich (2005). Smooth backfitting in practice. Journal of the Royal Statistical Society: Series B (Statistical Methodology) *67*(1), 43–61.

Panayotou, T. (1993). *Empirical tests and policy analysis of environmental degradation at different stages of economic development.* ILO.

Pareto, V. (1991). *The rise and fall of elites: an application of theoretical sociology.* Transaction Pub.

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. Econometric Theory, 597–625.

Perman, R. and D. Stern (2003). Evidence from panel unit root and cointegration tests that the environmental kuznets curve does not exist. Australian Journal of Agricultural and Resource Economics *47*(3), 325–347.

Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica *74*(4), 967–1012.

Pesaran, M. (2007). A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics *22*(2), 265–312.

Popp, D. (2005). Uncertain r&d and the porter hypothesis. *The BE Journal of Economic Analysis & Policy* (1).

Popp, D. (2006). International innovation and diffusion of air pollution control technologies: the effects of no< i> x</i> and so< sub> 2</sub> regulation in the us, japan, and germany. *Journal of Environmental Economics and Management 51*(1), 46–71.

Popp, D. and R. Newell (2012). Where does energy r&d come from? examining crowding out from energy r&d. *Energy Economics*.

Popp, D., R. Newell, and A. Jaffe (2010). Energy, the environment, and technological change. *Handbook of the Economics of Innovation 2*, 873–937.

Porter, M. and C. Van der Linde (1995). Toward a new conception of the environment-competitiveness relationship. *The Journal of Economic Perspectives 9*(4), 97–118.

Riahi, K., A. Grübler, and N. Nakicenovic (2007). Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change 74*(7), 887–935.

Ricci, F. (2007). Channels of transmission of environmental policy to economic growth: A survey of the theory. *Ecological Economics 60*(4), 688–699.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. Econometric Theory *7*(1), 1–21.

Schienle, M. (2011). Nonparametric nonstationary regression with many covariates. Technical report, SFB 649 discussion paper.

Schmalensee, R., T. M. Stoker, and R. A. Judson (1998). World carbon dioxide emissions: 1950-2050. Review of Economics and Statistics *80*(1), 15–27.

Selden, T. and D. Song (1994). Environmental quality and development: is there a kuznets curve for air pollution emissions? Journal of Environmental Economics and Management *27*(2), 147–162.

Sen, S., B. Melenberg, and H. Vollebergh (2014a). Appendix to the environmental kuznets curve: Identifying the nonlinear nonstationary scale effect. Technical report, Tilburg University.

Sen, S., B. Melenberg, and H. Vollebergh (2014b). The environmental Kuznets curve: Identifying the nonlinear nonstationary scale effects.

Shafik, N. and S. Bandyopadhyay (1992). *Economic growth and environmental quality: time-series and cross-country evidence*, Volume 904. World Bank Publications.

Simpson, R., R. Bradford, et al. (1996). Taxing variable cost: Environmental regulation as industrial policy. *Journal of Environmental Economics and Management 30*(3), 282–300.

Skocpol, T. (1979). *States and social revolutions: A comparative analysis of France, Russia, and China*. Cambridge Univ Pr.

Smith, S. J. and T. Wigley (2006). Multi-gas forcing stabilization with minicam. *Energy Journal 27.*

Stern, D. (1998). Progress on the environmental kuznets curve? Environment and Development Economics *3*(2), 173–196.

Stern, D. (2004). The rise and fall of the environmental kuznets curve. World development *32*(8), 1419–1439.

Taskin, F. and O. Zaim (2000). Searching for a kuznets curve in environmental efficiency using kernel estimation. Economics Letters *68*, 217–223.

Taylor, M. S. and B. R. Copeland (2004). Trade, growth, and the environment.

van Vuuren, D. P., M. G. Den Elzen, P. L. Lucas, B. Eickhout, B. J. Strengers, B. van Ruijven, S. Wonink, and R. van Houdt (2007). Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. *Climatic Change 81*(2), 119–159.

Verbeek, M. (2004). A guide to modern econometrics. southern gate, chichester, west sussex, england hoboken.

Vollebergh, H., B. Melenberg, and E. Dijkgraaf (2009). Identifying reduced-form relations with panel data: The case of pollution and income. Journal of Environmental Economics and Management *58*(1), 27–42.

Vries, F. and C. Withagen (2005). Innovation and environmental stringency: The case of sulfur dioxide abatement. Discussion Papers/CentER for Economic Research *2005*, 1–34.

Wagner, M. (2008). The carbon kuznets curve: A cloudy picture emitted by bad econometrics? Resource and Energy Economics *30*(3), 388–408.

Wang, Q. and P. Phillips (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory *25*(3), 710.

Westerlund, J. (2005). Data dependent endogeneity correction in cointegrated panels. Oxford Bulletin of Economics and Statistics *67*(5), 691–705.

Westerlund, J. (2007). Testing for error correction in panel data. Oxford Bulletin of Economics and Statistics *69*(6), 709–748.

Wise, M., K. Calvin, A. Thomson, L. Clarke, B. Bond-Lamberty, R. Sands, S. J. Smith, A. Janetos, and J. Edmonds (2009). Implications of limiting co2 concentrations for land use and energy. *Science 324*(5931), 1183–1186.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Xepapadeas, A. and A. de Zeeuw (1999). Environmental policy and competitiveness: The porter hypothesis and the composition of capital. *Journal of Environmental Economics and Management 37*(2), 165–182.